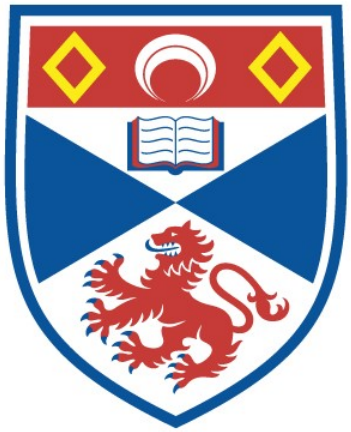
METHODS IN SPATIALLY EXPLICIT CAPTURE-RECAPTURE

Ben C. Stevenson

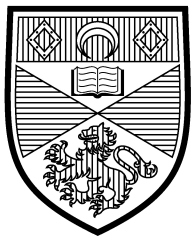
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2016

METHODS IN SPATIALLY EXPLICIT CAPTURE-RECAPTURE

Ben C. Stevenson



University of
St Andrews

A thesis submitted for the degree of
Doctor of Philosophy
at the University of St Andrews

23 March 2016

Candidate's declaration

I, Ben C. Stevenson, hereby certify that this thesis, which is approximately 53 000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in June 2012 and as a candidate for the degree of PhD also in June 2012; the higher study for which this is a record was carried out in the University of St Andrews between 2012 and 2016.

Date: 23/03/2016 Signature of candidate:

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 23/03/2016 Signature of supervisor:

Permission for publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of two years on the following grounds:

- Publication would preclude future publication.

Electronic copy

Embargo on all of print copy for a period of two years on the following grounds:

- Publication would preclude future publication.

Supporting statement for electronic embargo request

The ability to publish work found within this thesis may be compromised if it is not embargoed.

Date: 23/03/2016

Signature of candidate: Signature of supervisor:

Acknowledgements

I am grateful to both the Centre for Research into Ecological and Environmental Modelling (CREEM) and the National Centre for Statistical Ecology (NCSE) for studentships that provided funding for my studies. The NCSE studentship was available through a joint grant from the Engineering and Physical Sciences Research Council and the Natural Environment Research Council (#EP/1000917/1). Additional thanks to Auckland Grammar School, New Zealand, for further support through the F. W. W. Rhodes Memorial Scholarship.

I was fortunate to obtain funding from a variety of other sources that allowed me to undertake various worldwide adventures, and these greatly enriched my PhD experience. In particular, I acknowledge Hans Skaug for funding to attend the 2013 ADMB developers' meeting in Reykjavík, The National Geographic Society for a grant (via principal investigator John Measey) that part-funded a research trip to Cape Town, the University of Auckland for part-funding a collaborative research visit to New Zealand, Monique MacKenzie for taking me along to teach workshops in South Africa and Ghana, and The American Statistical Association for partial funding to attend the 2015 Joint Statistical Meetings in Seattle.

Despite ecological statistics being an interdisciplinary subject area, my previous training has been anything but. I thank those people who have put up with my general ignorance of the biological sciences—in particular John Measey and Res Altwegg, whose collaboration contributed strongly to the content found in Chapter 3. I am particularly obliged to John for proofreading passages from this thesis.

I am hugely indebted to Rachel Fewster. First, for being a fantastic collaborator (of the statistical kind); of all statisticians I have met she must be one of the finest at explaining her (usually complex) statistical ideas in an intuitive, understandable way. Furthermore, a research retreat she organised (and funded through a Royal Society of New Zealand Marsden grant) in the Scottish highlands has surely been one of the most productive weeks of research over the course of my studies. Much of the work found in Chapter 6 is due to this collaboration. Second, she was hugely influential in my decision to come to St Andrews to pursue studies towards a PhD in the first place. Despite never having properly met, in 2011 I organised a meeting with her enquire about the University of St Andrews as a possible candidate institution for further study. I left that meeting with the strongest recommendation to apply, and an excellent suggestion for a potential supervisor. I was blown away by the effort she had made into determining what I—a virtual stranger—was like (not only as a researcher, but as a person in general) in order to determine what would be the 'best fit' for me at St Andrews. The subsequent decision of where I was to pursue doctoral studies was not a difficult one to make. I recall that—after warning me about the

weather—she told me I would thoroughly enjoy working at CREEM.

And right she was. CREEM has provided the perfect working environment for my PhD studies: it has been intellectually stimulating and engaging place to study, and comprises the friendliest group of people I have ever had the pleasure of working with. I cannot name everyone personally, but I would like to acknowledge, in general, all who make it the place that it is. Thanks for the imparting your knowledge and advice (from some of the best minds in the field no less); thanks for the general chit-chat at coffee, the typically British small talk about how terrible the weather is, has been, and ever will be; and, of course, thanks for the cake.

I would like to specifically mention a few people who have had a particular influence on me during my time in St Andrews. I extend my gratitude to the following people: To Darren—former officemate—for his friendship; for his sense of humour, which made Room 115 such an enjoyable place to work; and for his contributions to discussions that (in particular) enhanced my understanding of SECR, C++, and the R package `Rcpp`. To Richard—replacement officemate and wise beyond his years—for all his perceptive advice: from which Linux distribution I should install (Xubuntu) and how powerful my laptop's processor should be (Intel Core i7), through to whether the rows in Table 3.2 should refer to parameters or confidence interval methods (the latter) and what type of cake I should bake at the weekend (upside-down mango). To Charlotte (and Watson) for the games, the adventures, and for keeping my chin up and a smile on my face—especially at times when there did not appear to be much to smile about. Finally, to Cassandra, whose love and support indirectly contributed to much of the work behind this thesis.

I make particular note of David Borchers, my supervisor, who has been outstanding from the very beginning. I had originally found it difficult to engage academics in distant lands for discussion about potential PhD studies at their institutions, often receiving generic replies about the application process and how difficult it might be to find appropriate funding. After sending David an initial e-mail expressing my interest in a PhD position at St Andrews, I awoke the next day to an enthusiastic, comprehensive reply (in excess of 1 500 words) detailing all the research areas in which he would be interested in providing supervision, and a range of potential funding options. His level of commitment to me and my work has never faltered. Time and again he has shown (possibly blind) faith in my abilities—for example, offering me the lead role of a project on the acoustic monitoring of frogs early on in my studies, and later sending me in his stead to present an invited talk at a leading international conference. He has allowed me the freedom to direct my research in whichever way I desired. All the same, he has always made himself available at the drop of

a hat for guidance and advice—not only about my research, but also on both personal and professional matters—despite his newfound professorial responsibilities. David has been everything I could have hoped for in not only a PhD supervisor, but also a mentor and guide to life in academia. These words do not do justice to the level of appreciation that I wish to convey.

Finally, a special thanks to my family back in New Zealand—Ollie, Kate, Nana, and particularly my parents, Helen and Craig. They have always been incredibly supportive of both my continued academic pursuits and my decision to undertake them on nearly the most distant landmass possible. Although our geographical separation may be vast, they have always been sure to maintain closeness in other (non-Euclidean) metrics. Thank you for always being an e-mail, a message, or a video call away, and for providing a constant source of love, support, and laughter.

Abstract

Capture-recapture (CR) methods are a ubiquitous means of estimating animal abundance from wildlife surveys. They rely on the detection and subsequent redetection of individuals over a number of sampling occasions. It is usually necessary for individuals to be recognised upon redetection. Spatially explicit capture-recapture (SECR) methods generalise those of CR by accounting for the locations at which each detection occurs. This allows spatial heterogeneity in detection probabilities to be accounted for: individuals with home-range centres near the detector array are more likely to be detected. They also permit estimation of animal density in addition to abundance.

One particular advantage of SECR methods is that they can be used when individuals are detected via the cues they produce—examples include birdsongs detected by microphones and whale surfacings detected by human observers. In such situations each cue may be detected by multiple detectors at different fixed locations. Redetections are then spatial (rather than temporal) in nature, and density can be estimated from a single survey occasion.

Existing methods, however, cannot generally be appropriately applied to the resulting cue-detection data without making assumptions that rarely hold. Additionally, they usually estimate cue density rather than animal density, which does not usually have the same biological importance. This thesis extends SECR methodology primarily for the appropriate estimation of animal density from cue-based SECR surveys. These extensions include (i) incorporation of auxiliary survey data into SECR estimators, (ii) appropriate point and variance estimators of animal density for a range of scenarios, and (iii) methods to account for both heterogeneity in detectability and cues that are directional in nature.

Moreover, a general class of methods is presented for the estimation of demographic parameters from wildlife surveys on which individuals cannot be recognised. These can variously be applied to CR and—potentially—SECR.

Contents

Chapter 1

Introduction

1.1 Estimating animal abundance and density

Animal abundance refers to the absolute number of individuals (usually of a particular species) that inhabit a particular site of interest. Animal density is the quotient of the abundance and the area of this site. Both variables are of great ecological importance: ecology has practical applications in a variety of fields, and for many of these it is necessary to estimate either abundance or density of animal species. Consider a conservation biologist who wishes to track the population size of a newly endangered species, a wildlife management consultant who is required to establish the efficacy of a recent pest eradication, and a policymaker looking to reduce the daily bag limit for recreational fishers in order to preserve the sustainability of a fishery. All require some method of gaining information about the density or abundance of at least one animal species.

In almost all cases an attempt to count each and every individual within the entire region of interest is infeasible. While these *census methods* are the only way of obtaining errorless measurements of abundance and density, they are also time consuming, they can be financially prohibitive, and are often practically or logistically impossible. In order to alleviate these burdens one may (i) only sample a subset of the region of interest, and (ii) forgo an exhaustive search, accepting that some individuals will be missed. While these practices decrease fieldwork effort, they introduce uncertainty: one must account for both the number of individuals outwith the searched area, and the number of individuals within the searched area that were missed. Neither can be known with certainty.

Estimating the probability of detecting individuals is central to (ii), above. A wealth of methods designed to estimate these detection probabilities (and therefore animal abundance or density) can be found in the statistical ecology literature. Two common approaches exist:

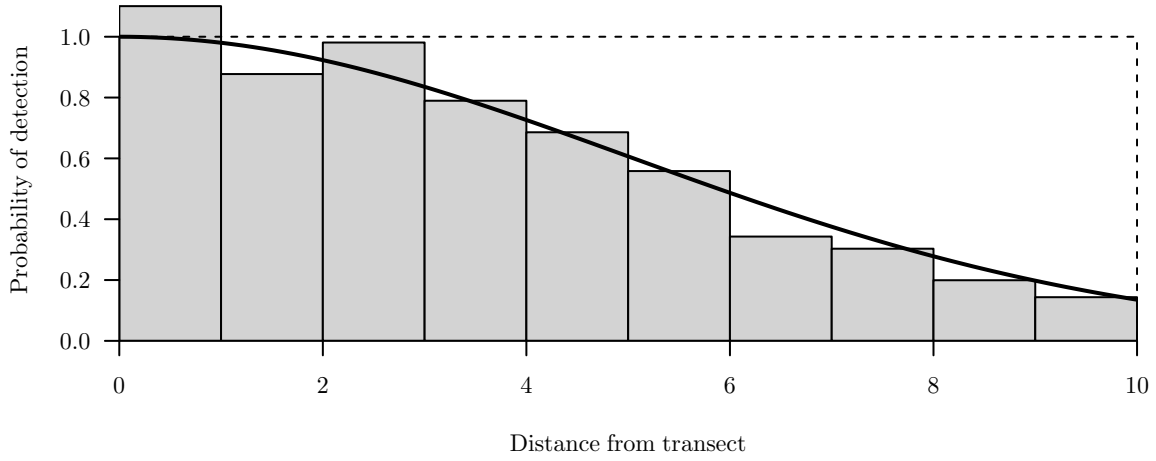


Figure 1.1 *A histogram of simulated distances to animals detected by observers walking along a line transect. Truncation distance was set at $w = 10$. The detection function used to simulate these data, $g(d)$, is overlain. The height of the histogram bars are scaled so that their collective area is equal to $\int_0^w g(x)\,\mathrm{d}x$, the area beneath the detection function.*

distance sampling (DS) and *capture-recapture* (CR). These are outlined below, along with *mark-recapture distance sampling* (MRDS), an approach that generalises DS via ideas from the CR literature.

1.1.1 Distance sampling

See Buckland et al. (2001) and Buckland et al. (2004) for general overviews of DS methodology. A DS survey is carried out by observers who either travel along line transects or are stationed at points. When an observer detects an individual (or group, if they occur in groups) they record its distance from either the line or the point. Any observations beyond some truncation distance w are ignored. If transects are placed at random, then the underlying distribution of all distances to individuals (detected or otherwise) within distance w is uniform on the interval $[0, w]$ for lines, or has probability density $f(d) \propto d, d \in [0, w]$, for points. However, as animals closer to the observer are more easily detected, there are a disproportionate number of *detected* individuals at shorter distances. The distribution of the observed distances allows estimation of a *detection function*—a mathematical function returning the probability of detecting an individual a given distance from the observer (Figure 1.1).

Obtaining an estimate of animal density involves the estimation of the proportion p, of all individuals within the truncation distance that were detected. As an example, for line-transect surveys this is given by the quotient of the areas beneath the true detection function

and the hypothetical detection function that arises from perfect detection (the dashed line in Figure 1.1), respectively. Let the total number of detections be n. An abundance estimate for the area searched (or *sampled*, i.e., the area within truncation distance w of the line or point) is $n/\widehat{p}$. Calculation of this area is straightforward: for a line transect of length l the sampled area is $2lw$, and for a point the sampled area is πw^2. As density is obtained by dividing abundance by the area searched, a density estimate $\widehat{D}$ is given by $n/(2l\widehat{p}w)$ for lines and $n/(\pi\widehat{p}w^2)$ for points.

The denominator in both cases is the estimated *effective sampling area* (ESA), here given by $\widehat{a}$. This is an estimate of the true ESA, a, and is the estiamted area of a region within which the expected number of individuals is n. For example, if a line transect survey is carried where the total area sampled is $100\,\mathrm{m}^2$ but the detection function is such that only half of all individuals within distance w were detected (i.e., $p = 0.5$), then the ESA is $0.5 \times 100 = 50\,\mathrm{m}^2$. Density estimates can equivalently be seen as the quotient of the number of detections made and the estimated ESA; that is, $\widehat{D} = n/\widehat{a}$. Also note that an estimate of animal abundance can be directly obtained by taking the product of the density estimate and the area of the site of interest.

1.1.2 Capture-recapture

CR methods (also known as 'mark-recapture' and 'capture-mark-recapture') have a long history in the field of abundance estimation, with the first such applications dating back to the estimation of the populations of seventeenth-century London (Graunt, 1662) and eighteenth-century France (Laplace, 1786). Here, only the estimation of closed populations (i.e., those without births, deaths, immigration, or emigration) is considered. Seber (1982) and Borchers, Buckland, and Zucchini (2002) provide overviews of various such CR approaches.

The underlying idea is straightforward: traps are laid out at fixed locations and these are used to physically capture animals. Individuals are captured on discrete sampling occasions in such a way that it is known with certainty when the same individual is captured again. This usually involves marking animals (e.g., by using a tag), but can alternatively rely on later recognition based on natural features (e.g., tigers' stripe patterns). The observed data are *capture histories*—one for each detected individual—denoting the sampling occasions on which the animals were detected. These data allow direct estimation of capture probabilities.

In general, animals need not be physically detained by traps, but can instead be detected in some other way (e.g., by visual detection from a camera). Therefore, terms such as

'detector' and 'detections' are henceforth used instead of traditional terms related to 'traps' and 'captures'[1].

As a simple example, consider a CR survey with just two discrete sampling occasions. Let the number of individuals detected on occasion i be n_i, the number of individuals captured on both occasions i and j be n_{ij}, and total abundance (the main parameter of interest) be N. Assume that capture events are independent across occasions and between individuals, and that the probability of capture p is the same for all individuals. A simple estimate for p is given by $\widehat{p} = n_{12}/n_1$; that is, the proportion of individuals marked on the first occasion that were detected again on the second occasion. An estimate for total population size can be calculated by scaling n_2 to account for undetected individuals; that is, $\widehat{N} = n_2/\widehat{p}$. This is the *Lincoln-Petersen method* of population abundance estimation; see Seber (1982) for further details. Alternative estimators are available for surveys with more than two capture occasions.

The assumption of a constant detection probability, p, across all individuals and over all capture occasions is seldom reasonable in practice. Otis, Burnham, White, and Anderson (1978) summarised various extensions to constant-probability models to allow for changes in detection probabilities (i) across time (denoted model M_t), (ii) across individuals (model M_h), and (iii) following an individual's first capture (e.g., due to trap-shy or trap-happy behaviour; model M_b). Combinations of the above can also be incorporated into the same model (e.g., model M_{tbh} to account for all three).

Unlike DS, CR methods do not have a spatial component. A consequence of this is that it is impossible to directly estimate density from the CR data alone—the sampled area is not known and model parameter estimates do not give rise to an estimate of the ESA. An estimate of abundance, $\widehat{N}$, can be problematic to interpret without an associated area: if the spatial range of the traps or detectors is small in comparison to the spatial range of the locations at which members of the population of interest live, then it is generally impossible to determine total population size—only those individuals that roam near the detectors are available for detection, and it is not possible to estimate the proportion of the population that this constitutes.

1.1.3 Mark-recapture distance sampling

A necessary assumption for many DS models is that detection is certain for animals situated directly on the transect (for line transect surveys) or at the observer location (for point transect surveys). MRDS methods (Borchers, Zucchini, & Fewster, 1998; Manly, McDonald,

[1] Although the term 'capture history' is still used for consistency with the CR literature.

& Garner, 1996) extend DS by including multiple, independent observers—each of whom can be considered a separate detection 'occasion'. This therefore uses ideas from both DS and CR: the CR component allows for estimation of the probability of detections at distance zero (i.e., animals either on the transect or at the point), while the DS component allows for the estimation of a detection function and a model for the distribution of observed distances.

The observers typically travel along the same transect (for line-transect surveys) or are situated on the same platform (for point-transect surveys; although in theory this need not necessarily be the case), and in both cases it is possible for the same animal to be detected by more than one observer. Detected individuals must be recognisable across observers in order to construct a capture history for each. A further important component is that the locations of detected animals on MRDS surveys are usually directly observed without error.

1.2 Spatially explicit capture-recapture

1.2.1 Overview

While model M_h (Otis et al., 1978) accounts for heterogeneity in detectability across individuals, CR methods do not directly model the obvious spatial component of these probabilities: individuals are more likely to be captured if they live and roam near the detectors. The relatively recent advent of methods that *do* incorporate spatial effects into detection probabilities fall into a class of model known as *spatially explicit capture-recapture* (SECR), although the term 'spatial capture-recapture' is also used. The first implementation of SECR models (Efford, 2004) relied on inverse prediction for parameter estimation. Maximum-likelihood (ML; Borchers & Efford, 2008) and Bayesian (Royle & Young, 2008) approaches followed shortly thereafter.

Regardless of the estimation method, SECR marries the spatial component of DS and the temporal nature of CR. A notable difference between SECR and CR is that CR models ignore the locations at which detections of individuals are made, while this information is retained in SECR. Capture histories for SECR therefore do not only include temporal detection information (i.e., the occasions on which an individual was detected), but also spatial detection information (i.e., where these detections occurred). This allows for the estimation of two additional model features: (i) the distribution of *activity centres* (sometimes called the 'home range centres') of the individuals, which can often be thought of as the animals' homes (e.g., their nests or burrows); and (ii) parameters of a detection function, returning the probability of detection given an activity centre distance d from the detector. Together these allow for the calculation of probabilities of detection by any detector at any time

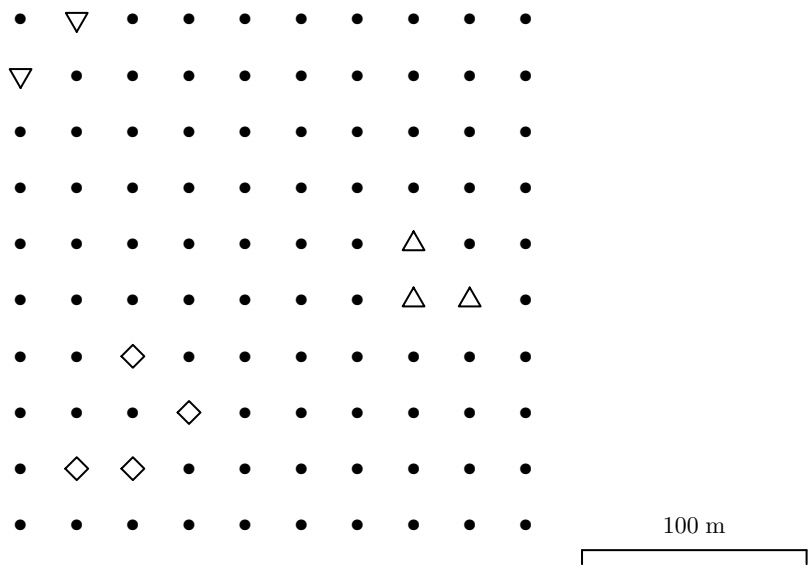


Figure 1.2 *A depiction of three hypothetical capture histories from an SECR survey carried out using a grid of* 100 *detectors. Detector locations are plotted. An open plotting character represents a location of a detector that detected one of the three individuals, where each plotting character type represents a different individual.*

for individuals with activity centres at any point in the survey region—thus quantifying interanimal heterogeneity in this regard. This also allows for the estimation of the ESA from the survey data alone, which is not otherwise possible using CR methods.

See Figure 1.2 for an example. From these data one can infer the approximate location of each individual's activity centre, as well as information about the detection function. In this case, it seems unlikely that a detector will detect an individual at a distance of more than 100 m from its activity centre. This, in turn, provides information about the sampled area (and thus animal density), which is usually unavailable in traditional CR analyses.

1.2.2 Spatial redetection

Initial implementations of SECR methodology relied on detection across discrete sampling occasions, as typically found in traditional CR surveys. Efford, Dawson, and Borchers (2009) made a further generalisation by noting that if redetections occur over a spatial dimension within the same sampling occasion (i.e., redetections of the same individual occur at different points in space rather than at different points in time), then animal abundance

and density parameters are estimable from only that sampling occasion alone.

This often occurs by virtue of individuals producing *cues*—stimuli emitted by animals that indicate their presence. Individuals may produce multiple cues, each of which may be detectable. Examples include songs produced by songbirds and surfacings of whales. In both cases, each cue may be simultaneously detected acoustically by multiple microphones (for the former) or visually by multiple human observers at fixed locations (for the latter). Conceptually, Figure 1.2 can be reconsidered with detectors as microphones, and open plotting characters representing locations of microphones that detected a particular cue produced by a particular individual.

Spatial redetection can also occur on surveys in which individuals do not emit cues that are detected by multiple detectors—sometimes this is instead induced by the movement of the animal itself. Common examples include camera-trap and hair-snag surveys; within the same sampling occasion an individual may be detected by cameras or snags at quite different locations. A further example that lies between the two above (cue and noncue) situations is that of scat surveys: each individual produces multiple cues (scat), however each is only ever once at a single location. The spatial range of detected scat produced by a particular individual is again due to its movement, rather than the propagation of the cue through space (cf. acoustic detection of animal vocalisations).

Surveys using spatial redetection methods hold a number of inherent advantages over those that require the physical detection of individuals over multiple occasions: (i) having only a single sampling occasion often results in cheaper and quicker surveys, particularly if the survey location is remote; (ii) spatial redetection methods tend to be *passive* in that individuals are not physically captured—passive detection is often less intrusive and safer (for the surveyors as well as the animals) than methods that require physical capture and tagging; (iii) these make SECR analyses possible for a larger number of species—for example, acoustic detection is useful for species that are visually cryptic and difficult to trap, but produce audible calls or can be detected by well-positioned camera traps.

There are certain similarities between the data collected for SECR with spatial redetection and those collected on MRDS surveys: in both cases, each animal can be detected by multiple detectors (perhaps human observers) that are situated at (potentially) different points in the survey region. An important distinction here is that the locations of detected individuals are assumed to be observed without error for MRDS, while this is not typically the case for SECR surveys. For the latter, only the locations of the *detectors* are known.

All methods discussed in this thesis refer to situations with cue-induced spatial redetection on surveys that have a single sampling occasion.

Table 1.1 *Examples of capture histories of the same hypothetical data with various definitions for the unit of detection. Each of the detector columns indicates which units were detected by the corresponding detector. A* 1 *indicates that the detector detected the unit, a* 0 *indicates that it did not. Tables (a) and (c) are cue-based capture histories, and so cue identification numbers are included. All tables include individual identification numbers.*

(a) *Cues as units of detection.*

		Detector			
Cue	ID	A	B	C	D
1	1	1	0	1	0
2	1	0	0	1	0
3	1	1	1	0	0
4	2	0	0	0	1
5	3	0	1	1	0
6	3	0	0	0	1

(b) *Individuals as units of detection.*

	Detector			
ID	A	B	C	D
1	1	1	1	0
2	0	0	0	1
3	0	1	1	1

(c) *First cues as units of detection.*

		Detector			
Cue	ID	A	B	C	D
1	1	1	0	1	0
4	2	0	0	0	1
5	3	0	1	1	0

1.2.3 Units of detection

For surveys on which individuals are detected from their cues, multiple cues from the same individual may be detected. Prior to analysis, the unit of detection must be appropriately defined. A natural choice is to select cues as the unit of detection, and so each detected *cue* contributes a capture history for analysis (e.g., Table 1.1a), giving rise to cue-based capture histories. Alternatively, individuals may be selected as the unit of detection, and so each detected *individual* contributes a single capture history for analysis (e.g., Table 1.1b), giving rise to individual-based capture histories. In this case, a detector is considered to have detected an individual if the detector detected at least one of its cues. Finally, the first detected cue from each individual may be used as the unit of detection (e.g., Table 1.1c), giving rise to first-cue-based capture histories. Dawson and Efford (2009) were the first to use this particular unit.

There are advantages and disadvantages associated with each of the three options—these are discussed throughout this thesis. Note that capture histories with either individuals or first cues as the unit of detection can only be created if individuals are recognisable from their cues (i.e., it is known with certainty whether or not any two detected cues were produced by the same individual). If not, only cue-based capture histories are available.

1.3 Statistical notation

Throughout Chapters 2–5, in order to prevent cumbersome notation, there is no explicit notational distinction made between a random variable and a fixed value in the support of its distribution—although any ambiguity will be resolved in writing (e.g., by stating 'the random variable x' or 'the observed value x'). Furthermore, $f(\cdot)$ is used to denote a probability density function (PDF; for a continuous random variable) or probability mass function (PMF; for a discrete random variable), while $F(\cdot)$ is used to denote a cumulative distribution function (CDF). Parameters corresponding to the distribution associated with the probability density functions (PDFs), probability mass functions (PMFs), CDFs, and other functions are typically separated from the arguments by a semicolon. Expectations and variances of random variables are denoted by the functions $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$, respectively.

Again, to preserve clarity, PDFs, PMFs, and CDFs do not typically explicitly indicate the corresponding random variable(s) to which they apply, as these should be clear from the arguments the function takes. An exception is made when ambiguity threatens, in which case a subscript following the function indicates the random variable it is associated with. For example, $f(x;\boldsymbol{\theta})$ may be used generically to give the value of the PDF of the continuous random variable x (associated with a distribution characterised by the parameter vector $\boldsymbol{\theta}$) at some fixed value x, while $f_x(0;\boldsymbol{\theta})$ is the evaluation of the same PDF at the value 0.

In general, bold lowercase font is used to denote a vector, while bold uppercase font is used to denote a matrix. Plain font denotes a scalar.

Chapter 6 does not involve the development of SECR methodology. The above does not apply, and notation specific to this work is described therein. A summary of almost all notation used throughout is available in two lists of notation (one for Chapters 1–5 and another for Chapter 6), found after the reference and acronym lists at the end of this thesis. These lists are comprehensive and should be consulted regularly—acronyms and notation are fully defined on first use, but not necessarily again.

1.4 Estimation

While the current surge in SECR methodological development is occurring under both the classical and Bayesian paradigms, throughout this thesis estimation is solely under the former (though this is not due to any ideological disposition held by the author).

This section formulates a basic likelihood, similar to that presented by Efford, Dawson, and Borchers (2009), for spatial-redetection SECR methodology. Aside from Chapter 6, the chapters that follow make use of the same (or similar) notation, and make various

modifications and additions to this likelihood. In particular, the development below does not apply for methods in which first cues are the unit of detection.

1.4.1 SECR notation

In total, m detectors of fixed and known location detect n units of detection (either individuals, cues, or first cues; see Table 1.1) over the course of a survey of duration t. The capture history for the ith unit is given by $\boldsymbol{\omega}_i$, where $\omega_{ij} = 1$ if it was detected by the jth detector, and $\omega_{ij} = 0$ otherwise. Capture histories for all detected units are contained in the matrix $\boldsymbol{\Omega}$, the ith row of which is given by $\boldsymbol{\omega}_i$.

Let the survey region be denoted by $\mathcal{A}$; this is the set of all habitable points at which a detected unit may be located. The location of the ith detected unit within the survey region is given by the Cartesian coordinates $\boldsymbol{x}_i = (x_{i1}, x_{i2}) \in \mathcal{A}$, and locations of all detected individuals are held in the matrix $\boldsymbol{X}$. At times, $\boldsymbol{x}$ is used generically to denote some point in the survey region that is not associated with any particular individual. Let $d_j(\boldsymbol{x})$ be the Euclidean distance between the location $\boldsymbol{x}$ and the jth detector. At times, d is used as a scalar to indicate some distance (e.g., as an argument to a detection function).

In general, SECR methods allow for estimation of a density surface, providing information not only about overall density, but also how it changes over the survey region. Here, however, only estimation of an unchanging density across $\mathcal{A}$ is considered, and this parameter is given by D. The definition of the units of detection provide the measurement unit of D: if individuals are detection units, then D is the density of individuals, while if cues are detection units, then D is the density of cues (i.e., cues emitted per unit area per time t). The vector $\boldsymbol{\gamma}$ contains parameters of the detection function, $g(\cdot)$, and let $\boldsymbol{\theta} = (D, \boldsymbol{\gamma})$.

1.4.2 The likelihood

Detection probabilities

The parametric function chosen to use as the detection function is achieved via a model selection approach. Two common choices are the *half-normal*,

$$g(x; \boldsymbol{\gamma}) = g_0 \exp\left(\frac{-x^2}{2\sigma^2}\right), \quad \boldsymbol{\gamma} = (g_0, \sigma), \tag{1.1}$$

and the *hazard-rate*,

$$g(x; \boldsymbol{\gamma}) = g_0 \left[1 - \exp\left(-\left(\frac{x}{\sigma}\right)^{-z}\right)\right], \quad \boldsymbol{\gamma} = (g_0, \sigma, z), \tag{1.2}$$

detection functions.

The probability of the jth detector detecting a unit located at $\boldsymbol{x}$ is $g(d_j(\boldsymbol{x});\boldsymbol{\gamma})$, by the definition of the detection function. Assuming independent detections across detectors, the probability that a unit at $\boldsymbol{x}$ is detected by at least one detector is

$$p(\boldsymbol{x};\boldsymbol{\gamma}) = 1 - \prod_{j=1}^{m}[1 - g(d_j(\boldsymbol{x});\boldsymbol{\gamma})], \tag{1.3}$$

where the product provides the probability of the unit evading detection. The function $p(\boldsymbol{x};\boldsymbol{\gamma})$ is referred to as the *detection probability surface* (DPS). As detectors are more likely to detect closer units, this surface is typically at its highest at points in the survey region close to the detectors, and approaches zero at locations further away.

Probability density of unit locations

Throughout methods in ecological statistics, it is often assumed that the locations of all individuals within the survey region (regardless of whether or not they were detected) are a realisation of a Poisson point process (e.g., Buckland, Oedekoven, & Borchers, 2016; Cormack & Jupp, 1991). This implies that their locations are independent of one another. Here such processes with homogeneous intensity D are considered.

The below only holds if individuals are the unit of detection. When this is the case, locations of *detected* individuals (and therefore units) are a realisation of a thinned Poisson point process, where the thinning is through the DPS (Equation (1.3)). The marginal PDF of the location of the ith detected individual is therefore

$$f(\boldsymbol{x}_i;\boldsymbol{\gamma}) = \frac{p(\boldsymbol{x}_i;\boldsymbol{\gamma})}{\int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{x}}, \tag{1.4}$$

and so the location of a detected individual is most likely to be at a point in $\mathcal{A}$ close to at least one detector, and unlikely to be at a point far from all detectors. Due to the aforementioned independence,

$$f(\boldsymbol{X}|n;\boldsymbol{\gamma}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i;\boldsymbol{\gamma}). \tag{1.5}$$

The denominator of Equation (1.4) gives the ESA:

$$a(\boldsymbol{\gamma}) = \int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{x}. \tag{1.6}$$

As detected individuals' locations are a realisation of a thinned Poisson point process, the total number of units detected is a random variable that has a Poisson distribution. Thus

$$n \sim \text{Poisson}(\mu),$$

where

$$\mu = \text{E}(n) = D \int_{\mathcal{A}} p(\boldsymbol{x}; \boldsymbol{\gamma}) \, \text{d}\boldsymbol{x},$$

and so

$$f(n; \boldsymbol{\theta}) = \frac{\mu^n \, e^{-\mu}}{n!}. \tag{1.7}$$

If cues are the unit of detection the above is unlikely to hold: locations of cues emitted by a particular individual are the same if it is stationary, and likely to be similar (thus not independent) if it is not. It follows that the joint PDF of the locations of detected units cannot be obtained via a product of the marginal densities, as per Equation (1.5). Furthermore—given that the number of individuals is a Poisson random variable—the distribution of the number of detected cues is unclear, and so Equation (1.7) no longer holds.

Probability mass of capture histories

The details of this section hold regardless of how the unit of detection is defined. Given the ith detected unit's location, the PMF of its capture history is

$$f(\boldsymbol{\omega}_i | \boldsymbol{x}_i; \boldsymbol{\gamma}) = \frac{\prod_{j=1}^{m} g(d_j(\boldsymbol{x}_i); \boldsymbol{\gamma})^{\omega_{ij}} \left[1 - g(d_j(\boldsymbol{x}_i); \boldsymbol{\gamma})\right]^{1-\omega_{ij}}}{p(\boldsymbol{x}_i; \boldsymbol{\gamma})}. \tag{1.8}$$

The numerator is a product across all detectors, where a detector contributes the probability of detection to the product if it detected the individual, or the probability of evasion if it did not. A capture history can only be observed if at least one detector makes a detection (i.e., observing $\boldsymbol{\omega}_i$ is conditional on $\sum_{j=1}^{m} \omega_{ij} > 0$) thus this product must be divided by $\Pr(\sum_{j=1}^{m} \boldsymbol{\omega}_i > 0) \equiv p(\boldsymbol{x}_i; \boldsymbol{\gamma})$ to account for the fact that a capture history of $\mathbf{0}_m$ cannot be observed.

Assuming independence between individuals gives the result

$$f(\boldsymbol{\Omega}|n, \boldsymbol{X}; \boldsymbol{\gamma}) = \prod_{i=1}^{n} f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma}). \tag{1.9}$$

The likelihood function

The likelihood is the joint density of the observed data, as a function of the model parameters; that is,

$$\begin{aligned} L(\boldsymbol{\theta}; n, \boldsymbol{\Omega}) &= f(n, \boldsymbol{\Omega}; \boldsymbol{\theta}) \\ &= f(n; \boldsymbol{\theta})\, f(\boldsymbol{\Omega}|n; \boldsymbol{\gamma}). \end{aligned} \tag{1.10}$$

The first term in the above product is available from Equation (1.7). For the second, there is obvious dependence between $\boldsymbol{\Omega}$ and $\boldsymbol{X}$: detectors are more likely to detect units they are close to, and less likely to detect those that are far away. Unit locations are unobserved, however, and so are modelled as latent variables. They are therefore incorporated into the PMF of $\boldsymbol{\Omega}$ via integration over the survey region:

$$\begin{aligned} f(\boldsymbol{\Omega}|n; \boldsymbol{\gamma}) &= \int_{\mathcal{A}^n} f(\boldsymbol{\Omega}, \boldsymbol{X}|n; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{X} \\ &= \int_{\mathcal{A}^n} f(\boldsymbol{\Omega}|n, \boldsymbol{X}; \boldsymbol{\gamma})\, f(\boldsymbol{X}|n; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{X}. \end{aligned} \tag{1.11}$$

In general this $2n$-dimensional integral is intractable, and at best its approximation is extremely computationally expensive. Assuming independence across units (both between their capture histories, conditional on location, and between their locations) allows for separability due to Equations (1.5) and (1.9), giving

$$\begin{aligned} f(\boldsymbol{\Omega}|n; \boldsymbol{\gamma}) &= \int_{\mathcal{A}^n} \prod_{i=1}^{n} [f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})]\, \mathrm{d}\boldsymbol{X} \\ &= \int_{\mathcal{A}} \cdots \int_{\mathcal{A}} \prod_{i=1}^{n} [f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})]\, \mathrm{d}\boldsymbol{x}_1 \cdots \mathrm{d}\boldsymbol{x}_n \\ &= \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}_i. \end{aligned} \tag{1.12}$$

The terms in the above integrand are available from Equations (1.8) and (1.4), respectively. Approximation of this product of n two-dimensional integrals is substantially more efficient. Regarding computation, the denominator of the first term (Equation (1.8)) and the

numerator of the second term (Equation (1.4)) in the integrand cancel. Furthermore, the numerator of the latter is a constant with respect to the integral and can be moved out of the integrand.

Estimation of $\boldsymbol{\theta}$ is via maximisation of the likelihood, and so

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; n, \boldsymbol{\Omega}). \tag{1.13}$$

1.5 Thesis overview

This thesis outlines a variety of methods aimed at improving animal abundance and density estimation from SECR survey data. With the exception of Chapter 6, the following chapters describe novel methods that involve generalisations and other modifications to the likelihood presented above.

A reference list immediately follows Chapter 6. Throughout this thesis, the first time a species is mentioned both common and binomial names are given; subsequently it is only referred to by the latter. A species list providing both common and binomial names appears after the reference list, and before the notation lists.

This section provides an overview of each chapter.

1.5.1 A unifying model for capture-recapture and distance sampling

With traditional SECR methodology information about an individual's location comes solely from the locations at which it is detected (see Figure 1.2). Surveys that collect data appropriate for analysis with SECR methods—particularly those that involve spatial redetection—are often capable of obtaining auxiliary data that may provide additional informative spatial information. For example, human observers acting as either visual or acoustic detectors may not only record whether or not they detected a particular individual, but they may also estimate a bearing or distance to the detected animal. Furthermore, acoustic detectors (such as microphones) may automatically measure strengths and precise times of arrival of detected signals. Detectors closer to an individual should expect to receive shorter estimated distances, and louder and sooner received signals.

In Chapter 2 methods for incorporating such auxiliary information into animal density estimators are described. It is shown that using such information can substantially improve estimator precision and also reduce bias. CR and DS are both special cases of this new model class, putting them at opposite ends of a spectrum of methods that vary in terms of the amount of spatial information employed.

Furthermore, it is noted that existing software does not generally allow for incorporation of these auxiliary data into SECR analyses. A new package for the R software environment for statistical computing and graphics (R Core Team, 2015) known as `admbsecr` is introduced. It is specifically designed for the analysis of data from SECR surveys on which auxiliary data has been collected. Chapter 2 therefore also describes this software package, and provides examples of its use. Indeed, many methods introduced in this thesis are implemented in the `admbsecr` package, and so its use is also shown throughout many of the following chapters.

1.5.2 Cue-based SECR methods

Section 1.4.2 provides comprehensive details of a likelihood for SECR methods when individuals are the units of detection. There are a number of issues that must be addressed for its use with cue-based capture histories. For example, (i) due to dependence between locations of cues emitted by the same individual, deriving the PMF of the number of detected units (Equation (1.7)), and the joint PDF of unit locations (Equation (1.5)) is not straightforward—these components remain unspecified for cue-based SECR methods; (ii) the simplification of an otherwise intractable likelihood function (Equation (1.12)) is only possible when the independence between locations holds; and (iii) the parameter D is cue density, but animal density is usually of interest.

The only published method for the analysis of cue detection data using SECR (Efford, Dawson, & Borchers, 2009) has involved a likelihood with the same features as that shown above. The approach is described implicitly assuming that no more than a single cue from each individual is detected; the assumption of independence between detected cues is therefore reasonable, and D corresponds to animal density. However, surveys are likely to record detections of multiple cues from each individual, and it is not clear how the method can be put into practice when this is the case. These shortcomings were not explicitly stated by Efford, Dawson, and Borchers (2009), and as a result a subsequent application of their method to acoustic detection data of the common minke whale *Balaenoptera acutorostrata* (Marques et al., 2012; Martin et al., 2013) involves the analysis of multiple cues per individual, despite the concerns addressed above.

In Chapter 3 the consequences of violating the assumption of independence between cue locations is explored, and methods for obtaining appropriate point and variance estimates of animal density from cue-based capture histories are described. These are applied to data from an acoustic survey of the Cape Peninsula moss frog *Arthroleptella lightfooti*.

1.5.3 First-cue SECR methods

The method of Efford, Dawson, and Borchers (2009) has also been applied to acoustic detection data of the ovenbird *Seiurus aurocapilla* (Dawson & Efford, 2009). In an attempt to meet the assumptions of the method only the first detected call from each detected individual was considered eligible for analysis. This can only be achieved if individuals are identifiable from their cues. However, this results in the estimation of a *cue*-based detection function; that is, $g(d;\boldsymbol{\gamma})$ and $p(\boldsymbol{x};\boldsymbol{\gamma})$ return the probabilities (at a single detector, and across an array of detectors, respectively) that a particular *cue* is detected. This is not appropriate when attempting to estimate animal density, as the probability of a single cue being detected is not equivalent to the probability of a particular animal being detected. Cues do not have a temporal component of detection as they occur at a (virtually) discrete point in time; however, individuals do have such a temporal component: the longer a survey, the more times an individual is likely to emit a cue—thus it is more detectable, and its detection probability is necessarily higher.

In Chapter 4 it is shown that estimating animal density using the method of Efford, Dawson, and Borchers (2009) by only including the first detected cue in the set of capture histories to analyse (e.g., Table 1.1c, and as per Dawson & Efford, 2009) is not appropriate, and that this practice can potentially result in substantially biased inference. Furthermore, a likelihood is derived, maximisation of which results in an animal density estimator with negligible bias that is shown to perform at least as well in all situations for first-cue capture histories (as per Dawson & Efford, 2009) as that proposed by Efford, Dawson, and Borchers (2009).

1.5.4 SECR methods for cue directionality and strength heterogeneity

The likelihood presented in Section 1.4.2 assumes that the detection probability of a cue at a particular detector only depends on the distance between their respective locations. Additionally, detections across the detectors are assumed to be independent, conditional on the cue's location (see Equation (1.8)). These features imply (i) that cues are equally detectable in all directions, and (ii) that all cues are equally detectable. This is consistent throughout the methods introduced in Chapters 3 and 4. These assumptions are not always realistic, particularly for acoustic surveys. Calls from some species are very directional, and are far more detectable in the direction they are emitted. For others, there may be marked heterogeneity in the strength a cue is emitted, so that some cues are more detectable than others.

In Chapter 5 methods are described that are able to cope with these features, estimating

either cue directionality (CD) or cue strength heterogeneity (CSH). They do not require the direction or cue strength to be observed—instead, they are able to estimate these features of the call. It is shown that estimators from these methods outperform those that ignore CD and CSH.

1.5.5 Uncertain identification in surveys of wildlife populations

Analysis of data collected from wildlife surveys often requires each detected animal to be identified and subsequently recognised at each redetection. This holds for SECR surveys, and indeed the methods presented in the aforementioned chapters all rely on the ability to ascertain which cues detected across an array were emitted by the same source.

Traditionally, CR surveys achieve identification by physically capturing and tagging individuals. With recent advancements in wildlife monitoring technology there is no longer the need to physically detain individuals—monitoring using cameras, microphones, and the like are gradually becoming commonplace. These allow the collection of more data than ever before despite a greatly reduced fieldwork burden. Identification approaches involve recognising unique features of individuals (such as scars, natural patterns, or acoustic signatures), or from genetic identification (from samples of hair or scat). However, these methods do not come with the same identification certainty as a securely fastened tag: similar physical or acoustic features (for visual or acoustic identification) and allele dropout (for genetic identification) introduce uncertainty into the identification of individual animals.

In Chapter 6 a method is developed for the estimation of whale density from an aerial survey on which planes fitted with high-definition cameras monitor a transect. Surfacing whales can be detected from the collected images. Density is not estimable if just a single plane is deployed, as the proportion of whales that were detected (and therefore those that remained undetected) is not known. There is potential for this proportion to be estimated if multiple planes are used: if individuals can be recognised across images collected by different planes the data collected can be analysed using straightforward CR methods, where each plane can be thought of as a capture occasion.

However, individuals cannot be recognised, and the method that is presented accounts for this identification uncertainty. It is one of the first members of a class recently coined as *trace-contrast* (TC) models. There is potential for the development of further related models of this class to provide a way of accounting for uncertain identification in a range of wildlife surveys, including SECR.

Chapter 2

A unifying model for capture-recapture and distance sampling

2.1 Introduction

As presented in Sections 1.1.1 and 1.1.2, one fundamental difference between traditional DS and CR methods is that DS incorporates the locations of detected animals (as distances from observers' locations), while CR does not. SECR methods sit in between: locations (e.g., of animals' activity centres, if traps are used as detectors) are not directly observed, but are nevertheless modelled as latent variables. Noisy information about the location associated with each detection unit is therefore held in its capture history—an individual is most likely to be close to the locations of the detectors that made a detection.

Remote detectors are often capable of collecting additional noisy information about locations associated with the detection units. Let $\boldsymbol{Y}$ contain all such information collected from all detections and $\boldsymbol{\psi}$ be a vector containing parameters required to model these data. Four types of such information are specifically dealt with in this chapter; these are (i) estimated bearings to detected animals, (ii) estimated distances to detected animals, (iii) received acoustic signal strengths and (iv) received acoustic signal times of arrival (TOAs).

Intuitively, the more information incorporated into statistical models the higher the precision of the resulting estimators. Reduced uncertainty about detection unit locations provides cleaner information about the distances between these locations and the detectors; this, in turn, allows the detection function parameters, $\boldsymbol{\gamma}$, to be estimated with more precision, and one would expect the reduced uncertainty to propagate through to the density

estimator.

A recent result also provides theoretical justification. Fewster and Jupp (2013) showed that incorporation of additional data sources into parameter estimators will result in confidence intervals (CIs) that are asymptotically narrower, despite the requirement for additional nuisance parameters. Thus, estimation via maximisation of a likelihood incorporating these auxiliary data,

$$(\widehat{D}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\psi}}) = \underset{D,\boldsymbol{\gamma},\boldsymbol{\psi}}{\arg\max}\, L(D, \boldsymbol{\gamma}, \boldsymbol{\psi}; n, \boldsymbol{\Omega}, \boldsymbol{Y}), \tag{2.1}$$

is preferable to that shown in Equation (1.13), as long as $\boldsymbol{Y}$ provides some information about the parameter of interest, D.

In this chapter, the likelihood shown above, $L(D, \boldsymbol{\gamma}, \boldsymbol{\psi}; n, \boldsymbol{\Omega}, \boldsymbol{Y})$, is derived, allowing estimation that incorporates general auxiliary information types (Section 2.2). For this likelihood individuals are defined as detection units, and so $\widehat{D}$ is the estimated animal density. It is shown that this likelihood is equivalent to those used for density estimation with MRDS and DS methods in the special case when $\boldsymbol{Y}$ provides perfect, noiseless information about detected animals' locations. Moreover, it is equivalent to those used for abundance estimation with standard CR methods when there is no spatial effect on detection probabilities. This thus unifies CR, SECR, and DS under a single model class.

Additionally, under a particular set of assumptions, the special case of modelling observed signal strengths reduces to the model proposed by Efford, Dawson, and Borchers (2009) for estimation of animal density from an array of acoustic detectors. The combination of estimated distances and bearings is particularly useful: the auxiliary information collected by observers making visual detections of individuals may relate to where the animal was thought to have been seen, and this location estimate can then be decomposed into bearing and distance estimates.

The ideas presented and the likelihood derived in Section 2.2 have recently been published (Borchers, Stevenson, Kidney, Thomas, & Marques, 2015) in the *Journal of the American Statistical Association*, March 2015. The author of this thesis contributed to this work primarily through the development of the R package `admbsecr`, which provides a software implementation of ML estimation for SECR models with incorporated auxiliary information. A description of this package is provided in Section 2.3. Further contributions were made by using this R package to estimate animal density from various sets of data, and by conducting a simulation study relevant to each; this allowed for an assessment of estimator performance and investigation of the effect due to auxiliary data incorporation. Details of these applications are found in Section 2.4, along with the corresponding simulation studies. Analyses carried out were on (i) acoustic detection data of *A. lightfooti*

collected by an array of six microphones, which recorded both the signal strength and TOA of each detected call; (ii) acoustic detection data of the northern yellow-cheeked gibbon *Nomascus annamensis* collected by three observers, who recorded the estimated bearing to each detected gibbon group; and (iii) visual detection of surfacing *B. acutorostrata* collected by two observers aboard a plane, who recorded estimated distances to each detected whale.

A manuscript detailing an in-depth analysis of the *N. annamensis* detection data that included further methodological development not presented here (Kidney et al., in submission) has been submitted to *PLoS ONE*.

This chapter chiefly focusses on Sections 2.3 and 2.4, as these specifically describe the work undertaken by the author of this thesis towards the publication that appeared in the *Journal of the American Statistical Association*.

2.2 Incorporation of auxiliary information into the SECR likelihood

From here, let $\boldsymbol{\theta}$ include the additional nuisance parameters $\boldsymbol{\psi}$, so that $\boldsymbol{\theta} = (D, \boldsymbol{\gamma}, \boldsymbol{\psi})$. Incorporating auxiliary information—and therefore also the parameter vector $\boldsymbol{\psi}$—into the basic SECR likelihood presented in Section 1.4.2 involves a simple extension to Equation (1.10):

$$\begin{aligned} L(\boldsymbol{\theta}; n, \boldsymbol{\Omega}, \boldsymbol{Y}) &= f(n, \boldsymbol{\Omega}, \boldsymbol{Y}; \boldsymbol{\theta}) \\ &= f(n; D, \boldsymbol{\gamma})\, f(\boldsymbol{\Omega}, \boldsymbol{Y} | n; \boldsymbol{\gamma}, \boldsymbol{\psi}). \end{aligned} \tag{2.2}$$

The first term remains as per Equation (1.7), and the second can be expanded to give

$$\begin{aligned} f(\boldsymbol{\Omega}, \boldsymbol{Y} | n; \boldsymbol{\gamma}, \boldsymbol{\psi}) &= \int_{\mathcal{A}^n} f(\boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{X} | n; \boldsymbol{\gamma}, \boldsymbol{\psi})\, \mathrm{d}\boldsymbol{X} \\ &= \int_{\mathcal{A}^n} f(\boldsymbol{\Omega}, \boldsymbol{Y} | n, \boldsymbol{X}; \boldsymbol{\gamma}, \boldsymbol{\psi})\, f(\boldsymbol{X} | n; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{X} \\ &= \int_{\mathcal{A}^n} f(\boldsymbol{Y} | n, \boldsymbol{\Omega}, \boldsymbol{X}; \boldsymbol{\psi})\, f(\boldsymbol{\Omega} | n, \boldsymbol{X}; \boldsymbol{\gamma})\, f(\boldsymbol{X} | n; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{X}. \end{aligned}$$

Assuming independence between individuals (as per Equation (1.12)) gives

$$f(\boldsymbol{\Omega}, \boldsymbol{Y} | n; \boldsymbol{\gamma}, \boldsymbol{\psi}) = \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{Y}_i | \boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i | \boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}_i. \tag{2.3}$$

Here, $\boldsymbol{Y}_i$ contains all observed auxiliary information for the ith detected animal, across all detectors. Furthermore, $\boldsymbol{y}_{ij}$ gives all observed auxiliary information for the ith detected animal at the jth detector. As more than one type of auxiliary information may have been collected, $\boldsymbol{y}_{ij}$ is itself a vector, and y_{ijk} is an element providing the kth type. Finally, let $\boldsymbol{y}_{i\cdot k}$ be a vector of the kth auxiliary information type from detections of the ith animal, across all detectors. Detectors that did not detect an individual are not able to collect auxiliary information about the detection, so y_{ijk} is only observed if $\omega_{ij} = 1$.

Each y_{ijk} is considered a realisation of a random variable, allowing for the modelling of measurement error in the collection of auxiliary information. For example, visual observers may record estimates of the locations of detected individuals (which are then decomposed into distance and bearing estimates); however, these are not simply assumed to be true, exact locations (as per DS methodology), but rather distributions are chosen to model the uncertainty in these distance and bearing estimates—parameters for these form $\boldsymbol{\psi}$. Likewise, upon collection of exact TOAs, the locations of individuals detected by three or more detectors can be determined analytically; however, even a slight imprecision in recorded times may result in these calculated locations being too far from the true locations for the model to provide suitable estimates. The framework presented in this chapter allows for the modelling of this measurement error and estimation of its variance.

The second and third terms in the integrand of Equation (1.12) remain equivalent to those presented in Equations (1.8) and (1.4), respectively; all that remains is to derive the first, $f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$.

In many cases it is appropriate to assume independence between information types, conditional on animal detection and location (i.e., between the vectors of random variables $\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i$ and $\boldsymbol{y}_{i\cdot k'}|\boldsymbol{\omega}_i, \boldsymbol{x}_i$, $k \neq k'$). This is an appropriate assumption for all four combinations of auxiliary types considered here: It is sensible to assume that estimated bearings do not depend on whether distances have been over- or underestimated, and sound signals received at louder amplitudes do not travel any faster or slower. This independence assumption gives

$$f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi}) = \prod_{k=1}^{q} f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi}), \tag{2.4}$$

where the number of observed auxiliary information types is given by q.

The form of the joint PDF $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$ varies depending on the auxiliary information type in question. Typically, there are two components that characterise its chosen distribution: the expectation of the observed auxiliary data, usually a function of its conditional location, $\boldsymbol{x}_i$, and a parameter measuring the error that it is subject to. Let $\mathrm{E}(\boldsymbol{y}_{i\cdot k}|\boldsymbol{x}_i; \boldsymbol{\psi})$

return a vector of expectations of the kth auxiliary data type across the m detectors, conditional on its location $\boldsymbol{x}_i$.

Below $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$ is defined for each of the four auxiliary information types considered here.

2.2.1 Estimated distances

For this auxiliary information type each detection made comes with a corresponding estimate of the distance between the animal and the detector that made the detection. Below this is assumed to be the kth auxiliary information type. While this framework allows for measurement error in the distance estimates, it is assumed that they are unbiased (i.e., that the expected estimate is equal to the true distance). This gives the result

$$\mathrm{E}(y_{ijk}|\boldsymbol{x}_i; \boldsymbol{\psi}) = d_j(\boldsymbol{x}_i). \tag{2.5}$$

Relaxation of this assumption is discussed in Section 2.5.

It is also assumed that—conditional on animal location—distance estimates are independent across detectors (i.e., the size of the error made by one detector does not affect that of another). Therefore,

$$f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi}) = \prod_{\{j:\omega_{ij}=1\}} f(y_{ijk}|\boldsymbol{x}_i; \boldsymbol{\psi}). \tag{2.6}$$

Estimated distances are continuous in nature and necessarily positive, and so a suitable distribution chosen to model these should have the positive real numbers as its support. Examples include the gamma and the log-normal distributions. The former is described below, but derivation of the latter (or any other sensible choice) is readily obtainable.

The parameterisation of the gamma distribution used here has a shape, α, and a rate, β, parameter; its PDF is

$$f(y; \alpha, \beta) = \frac{\beta^\alpha \, y^{\alpha-1} \, e^{-\beta y}}{\Gamma(\alpha)}.$$

Its expectation is α/β, and recall that this is also given by Equation (2.5). Thus, for this application, $\beta = \alpha/d_j(\boldsymbol{x})$, and so

$$f(y_{ijk}|\boldsymbol{x}_i; \boldsymbol{\psi}) = \frac{\alpha^\alpha \, y_{ijk}^{\alpha-1} \, \exp(-\alpha \, y_{ijk}/d_j(\boldsymbol{x}_i))}{d_j(\boldsymbol{x})^\alpha \, \Gamma(\alpha)}, \tag{2.7}$$

where $\alpha \in \boldsymbol{\psi}$ ($\alpha \equiv \boldsymbol{\psi}$ if estimated distances are the only auxiliary information type). The

parameter α can be thought of as a parameter controlling the variance of the measurement uncertainty in distance estimates.

2.2.2 Estimated bearings

For this auxiliary information type each detection made comes with a corresponding estimate of the bearing from the detector that made the detection to the animal detected, measured in radians from some reference bearing (e.g., north). Below this is assumed to be the kth auxiliary information type.

It is again assumed that estimates are unbiased and independent across detectors. Due to unbiasedness

$$\mathrm{E}(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi}) = b_j(\boldsymbol{x}_i), \tag{2.8}$$

where $b_j(\boldsymbol{x}_i)$ gives the bearing from the jth detector to the conditioned location of the ith detected individual. Due to independence Equation (2.6) applies once more.

Estimated bearings are given in the interval $[0, 2\pi)$, and so a suitable circular distribution with this support should be chosen to model the estimated bearings; examples include the von Mises, wrapped-normal, and wrapped-Cauchy distributions. Here the former is described—but, once again, derivations using the latter two are readily obtainable.

The von Mises distribution has two parameters: its expectation (here given by $b_j(\boldsymbol{x}_i)$, Equation (2.8)), and κ, controlling its variance. This gives the PDF

$$f(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi}) = \frac{\exp(\kappa\,\cos(y_{ijk} - b_j(\boldsymbol{x}_i)))}{2\pi\, I_0(\kappa)} \tag{2.9}$$

where $\kappa \in \boldsymbol{\psi}$ ($\kappa \equiv \boldsymbol{\psi}$ if estimated bearings are the only auxiliary information type) and $I_0(\cdot)$ is the Bessel function of order 0. The parameter κ can be thought of as a parameter controlling the variance of the measurement uncertainty in bearing estimates.

2.2.3 Received signal strengths

For this auxiliary information type each detection made comes with a corresponding strength of the received signal, for example, measured in dB. Below this is assumed to be the kth auxiliary information type.

Signal strengths are inextricably linked with detection probabilities: a strong received acoustic signal is likely to be detected, but a weak received signal may be undetectable over any ambient background noise (Efford, Dawson, & Borchers, 2009). Received signal strengths are therefore modelled slightly differently to bearing and distance estimates, and

are incorporated directly into the detection function. The additional parameters used to model the signal strength information are therefore detection function parameters; they therefore do not contribute to the parameter vector $\boldsymbol{\psi}$ and instead make up $\boldsymbol{\gamma}$. Here the *signal strength* detection function is used in place of half normal (Equation 1.1), hazard rate (Equation 1.2), or similar alternatives.

This is achieved by specifying some threshold cutoff value, c, set at a level higher than any background noise so that a received signal stronger than c will certainly be detected. Any signals received at a level lower than c are ignored and considered nondetections; that is, $\omega_{ij} = 1$ if $y_{ijk} \geq c$ and $\omega_{ij} = 0$ otherwise. Let y^*_{ijk} be the strength of the ith individual's signal upon reaching the jth detector, regardless of whether or not it corresponds to a detection. Thus y^*_{ijk} is not necessarily observed if the signal was too weak to detect at all over any background noise, but $y^*_{ijk} = y_{ijk}$ if $y^*_{ijk} \geq c$.

Received signals are typically weaker at detectors further from the signal's source. It is assumed that the expected received signal strength decreases linearly (on the scale of some link function) with detector-to-source distance—that is, $\mathrm{E}(y^*_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi}) = h^{-1}(\beta_0 - \beta_1\, d_j(\boldsymbol{x}_i))$. Here $h^{-1}(\cdot)$ corresponds to the inverse of the link, $h(\cdot)$, with common choices being either the identity or exponential functions (corresponding to identity and log links, respectively). Assuming Gaussian measurement error for received signal strengths,

$$y^*_{ijk}|\boldsymbol{x}_i \sim \mathrm{N}(\mathrm{E}(y^*_{ijk}|\boldsymbol{x}_i;\boldsymbol{\gamma}), \sigma_s). \tag{2.10}$$

Recall that an individual is detected if its received signal strength exceeds c, and so, conditional on an individual being located distance d from a detector, this occurs with probability

$$\begin{aligned} g(d;\boldsymbol{\gamma},c) &= 1 - F_{y|d}(c|d;\boldsymbol{\gamma}) \\ &= 1 - \Phi\left(\frac{c - h^{-1}(\beta_0 - \beta_1 d)}{\sigma_s}\right), \end{aligned} \tag{2.11}$$

giving rise to the signal strength detection function. Here $\Phi(\cdot)$ is the CDF of the standard normal distribution. For the calculation of the likelihood, this detection function is then used to calculate the DPS (Equation (1.3)) and the probability mass of capture histories (Equation (1.8)).

Due to the truncation of all received signal strengths weaker than c, each observed signal strength is a realisation of a truncated normal random variable, and so has PDF

$$f(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\gamma},c) = \frac{f(y^*_{ijk}|\boldsymbol{x}_i;\boldsymbol{\gamma})}{1 - F_{y^*|\boldsymbol{x}}(c|\boldsymbol{x}_i;\boldsymbol{\gamma})}$$

$$= \frac{\phi([y_{ijk} - h^{-1}(\beta_0 - \beta_1\, d_j(\boldsymbol{x}_i))]/\sigma_s)}{\sigma_s\{1 - \Phi([c - h^{-1}(\beta_0 - \beta_1\, d_j(\boldsymbol{x}_i))]/\sigma_s)\}}. \tag{2.12}$$

The function $\phi(\cdot)$ is the PDF of the standard normal distribution. The joint distribution across all detectors, $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\gamma})$, is available via an assumption of independence (Equation (2.6)).

An alternative formulation based on the physical properties of the spherical spreading of sound energy, gives rise to the *spherical-spreading signal strength detection function* (Dawson & Efford, 2009). It replaces

$$\mathrm{E}(y^*_{ijk}|\boldsymbol{x}_i; \boldsymbol{\gamma}) = h^{-1}(\beta_0 - \beta_1\, d_j(\boldsymbol{x}_i)),$$

above, with

$$\mathrm{E}(y^*_{ijk}|\boldsymbol{x}_i; \boldsymbol{\gamma}) = \beta_0 - 10\, \log_{10}(d_j(\boldsymbol{x}_i)^2) + \beta_1[d_j(\boldsymbol{x}_i) - 1] \tag{2.13}$$

The threshold detection function

In some situations signal strengths may not be observed, but it may be desirable to fit a detection function with the same functional form as the signal strength detection function. However, in this case, it is not possible to make direct inference about source signal strength (β_0), signal strength loss (β_1), or associated measurement error (σ_s).

Using the identity function for $h(\cdot)$, the signal strength detection function can be rewritten as

$$\begin{aligned} g(d; \boldsymbol{\gamma}, c) &= 1 - \Phi\left(\frac{c - (\beta_0 - \beta_1 d)}{\sigma_s}\right) \\ &= 1 - \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{c - (\beta_0 - \beta_1 d)}{\sqrt{2\sigma_s^2}}\right)\right] \\ &= \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{c - \beta_0}{\sqrt{2\sigma_s^2}} + \frac{\beta_1}{\sqrt{2\sigma_s^2}} d\right), \end{aligned}$$

where $\mathrm{erf}(\cdot)$ is the error function. Substituting $\nu = (\beta_0 - c)/\sqrt{2\sigma_s^2}$ and $\tau = \sqrt{2\sigma_s^2}/\beta_1$ gives

$$g(d; \boldsymbol{\gamma}) = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{d}{\tau} - \nu\right),$$

the *threshold detection function*. The parameters ν and τ are known as the *shape* and *scale* parameters, respectively. Subsequent to the formulation of this detection function it was discovered that it is a reparameterisation of the so-called binary signal strength detection

function, defined by Efford, Dawson, and Borchers (2009) as

$$g(d;\boldsymbol{\gamma}) = 1 - \Phi(-(\alpha + \beta d)).$$

If $h(\cdot)$ is the $\log(\cdot)$ function (and so $h^{-1}(\cdot)$ is $\exp(\cdot)$), then rearrangement of Equation (2.11) gives

$$\begin{aligned} g(d;\boldsymbol{\gamma},c) &= 1 - \Phi\left(\frac{c - \exp(\beta_0 - \beta_1 d)}{\sigma_s}\right) \\ &= 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{c - \exp(\beta_0 - \beta_1 d)}{\sqrt{2\sigma_s^2}}\right)\right] \\ &= 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{c}{\sqrt{2\sigma_s^2}} - \frac{\exp(\beta_0 - \beta_1 d)}{\exp(\log(\sqrt{2\sigma_s^2}))}\right)\right] \\ &= 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{c}{\sqrt{2\sigma_s^2}} - \exp\left(\beta_0 - \beta_1 d - \log\left(\sqrt{2\sigma_s^2}\right)\right)\right)\right], \end{aligned}$$

and the *log-link threshold detection function* is derived following substitution of $\nu_1 = c/\sqrt{2\sigma_s^2}$, $\nu_2 = \log(\sqrt{2\sigma_s^2}) - \beta_0$, and $\tau = -1/\beta_1$:

$$g(d;\boldsymbol{\gamma}) = 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\nu_1 - \exp\left(\frac{d}{\tau} - \nu_2\right)\right)\right].$$

At present there has been no development of threshold detection functions for any other link $h(\cdot)$, nor for spherical spreading.

2.2.4 Received times of arrival

For this auxiliary information type each detection made comes with a corresponding precise time that the detected signal arrived at the detector. These data are considered to be measured in seconds from the beginning of the survey. Below this is assumed to be the kth auxiliary information type.

Received TOAs are again handled slightly differently to estimated distances or bearings, as the marginal distribution $f(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi})$ is not informative about the location $\boldsymbol{x}_i$. For example, the recorded time y_{ijk} could be due to the signal being produced either at the same time it was detected (by virtue of the detected individual being at the same location as the detector) or a full second prior to the time it was detected (by virtue of the individual being a distance of s m from the detector location; where s is the speed of sound in metres per second). It is the *differences* between the TOAs across detectors that provides the

information—an individual is most likely to be closest to the detector that first received the signal. It is the joint distribution of times of arrival across all detectors, $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$, that must be derived instead of the marginal PDFs.

This can be achieved by integrating over each cue's unknown emission time, considering it as a latent variable. Borchers et al. (2015; in the supplementary materials) showed that this gives the closed-form expression

$$f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi}) = \frac{(2\pi\,\sigma_t^2)^{(1-z_i)/2}}{2t\,\sqrt{z_i}} \exp\left(\sum_{\{j:\omega_{ij}=1\}} \frac{[\delta(y_{ijk}, \boldsymbol{x}_i; s) - \bar{\delta}(\boldsymbol{y}_{i\cdot k}, \boldsymbol{x}_i; s)]^2}{-2\sigma_t^2}\right). \quad (2.14)$$

The function $\delta(y_{ijk}, \boldsymbol{x}_i; s)$ is the expected time of cue emission given its location $\boldsymbol{x}_i$ and time of detection y_{ijk} at the jth detector; that is, $\delta(y_{ijk}, \boldsymbol{x}_i; s) = y_{ijk} - d_j(\boldsymbol{x}_i)/s$. Furthermore, $\bar{\delta}(\boldsymbol{y}_{i\cdot k}, \boldsymbol{x}_i; s)$ is the average across all detectors; that is, $\bar{\delta}(\boldsymbol{y}_{i\cdot k}, \boldsymbol{x}_i; s) = \sum_{j=1}^{m} \delta(y_{ijk}, \boldsymbol{x}_i; s)/z_i$. Finally, z_i is the number of detectors that detected the ith individual, and so $z_i = \sum_{j=1}^{m} \omega_{ij}$.

TOA is only informative about an individual's location if it is detected by more than one detector (i.e., if $z_i = \sum_{j=1}^{m} \omega_{ij} > 1$) as otherwise there are no differences to calculate. This is reflected in Equation (2.14) in that, if $z_i = 1$, then $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$ reduces to a constant that is not affected by $\boldsymbol{\theta}$. Thus the TOA of the single detection made is not informative about any parameters.

2.2.5 DS and MRDS

Recall that animal locations are directly observed on DS and MRDS surveys. These can be considered auxiliary data, where $\boldsymbol{Y}_i$ simply gives the observed location of the ith detected individual; however, in this case, the auxiliary data are *not* subject to measurement error. Once $\boldsymbol{Y}$ is considered to give errorless location data the typical likelihoods for line- and point-transect DS and MRDS can be derived from the likelihood shown in Equation (2.3); see Borchers et al. (2015; in particular, the supplementary information) for further detail. These can be evaluated without integration over $\boldsymbol{X}$, which is unsurprising as detected animal locations are now no longer latent.

2.2.6 The model continuum

A family of models therefore exists along a continuum: at one end there are CR methods, where spatial information is not explicitly incorporated; next lies traditional SECR approaches, where animal locations are accounted for, but the only available information about these comes from the locations of the detectors; SECR models that incorporate auxil-

iary data allow the use of additional spatial information; finally, DS and MRDS approaches make use of errorless information about animal locations. All are unified under the model class described above as special cases of the derived likelihood.

2.3 The admbsecr package

The secr R package (Efford, 2015) provides the most general, up-to-date, widely used software implementation of SECR methodology. However, aside from received signal strengths, it does not support the use of the auxiliary data described above. The admbsecr package (Stevenson & Borchers, 2015) aims to fill this methodological gap in the existing software, and its main features are described here. Further methodology is developed in the chapters of this thesis that follow, all of which (aside from what is presented in Chapter 6) have implementations in this R package; these particular features of the package are described in the relevant chapters.

The admbsecr package comes bundled with an executable that is used to fit SECR models, and so it is not available on the Comprehensive R Archive Network (CRAN). Packages that include such binary files are not permitted for submission. For installation instructions visit https://github.com/b-steve/admbsecr. For further support see the package manual, as all package functions are fully documented. Additionally, there are two vignettes providing further information (i) about the required structure for SECR detection data, and (ii) specific to fitting models to acoustic SECR data, each accessed by running

```
vignette("data-structure")
vignette("acoustic-secr")
```

after loading the package, respectively.

The admbsecr package comes with two exported data sets—one of which, held in the object example, will be used here to demonstrate various aspects of the package here. The example data are simulated detections and observed auxiliary information (covering all types listed above) across an array of six detectors.

2.3.1 Model fitting

Fitting an SECR model that incorporates auxiliary information can be achieved using the admbsecr() function. It has a variety of arguments, many of which are discussed in this section, along with computational details associated with the estimation procedure.

Data input and model specification

The first mandatory argument of the `admbsecr()` function is `capt`, which contains the capture information (including auxiliary information) for each detected unit. This must be a list containing named components. One component must be named `capt`, containing the capture histories, $\boldsymbol{\Omega}$, as a matrix with n rows and m columns. The jth element of the ith row should provide ω_{ij}. The remaining components provide the auxiliary information, where names `dist`, `bearing`, `ss`, and `toa` correspond to estimated distance, estimated bearing, received signal strength, and received TOA information, respectively. For further detail see the aforementioned vignette on required data structures. For information about conversion of a data object suitable for use with the `secr` package see Section 2.3.6.

The second mandatory argument is `traps`, simply constituting a matrix with m rows and two columns, where the jth row provides the Cartesian coordinates of the jth detector.

The modelled detection function can be set using the `detfn` argument. This is done by specifying a character string, where `"hn"`, `"hr"`, `"th"`, and `"lth"` correspond to the half-normal, hazard-rate, threshold, and log-link threshold functional forms, respectively. If signal strength information is included in `capt`, then a signal strength detection function is specified automatically.

Further options for signal strength models can be set using the argument `ss.opts`. This takes a list with a number of possible named components. Examples include (i) `cutoff`: specifying the threshold value c; and (ii) `ss.link`: specifying the link function, $h(\cdot)$, to use for the signal strength detection function (Equation (2.11)). Possible choices for the latter include `"identity"` and `"log"`. Alternatively, `"ss.link"` can be set to `"spherical"` to specify the spherical-spreading signal strength detection function.

Optimisation

Parameter estimates for $\boldsymbol{\theta}$ are obtained via maximisation of the log-likelihood; that is, the log of Equation (2.2). The optimisation procedure is carried out by an executable generated by AD model builder (ADMB).

ADMB (Fournier et al., 2012) is an open-source software package that provides a powerful, efficient optimiser, well suited to the fitting of complex statistical models. The user must provide a template (`.tpl`) file that specifies (i) the structure of the data, held in a separate data (`.dat`) file; (ii) the parameters of the model and the structure of their start values, held in a separate parameter initial values (`.pin`) file; and (iii) the model's negative log-likelihood, written in a language based on (and similar to) C++. ADMB can then be used to convert this template into true C++ code, which is then compiled to create an ex-

ecutable. Running this executable (and specifying the relevant `.dat` and `.pin` files) results in the fitting of the specified model to the data at hand by ML. Optimisation is carried out numerically using a quasi-Newton method.

Newton's method is an optimisation algorithm that, at each step, relies on a local quadratic approximation of the log-likelihood surface via a second-order Taylor-series expansion. This quadratic approximation is straightforward to maximise, and a step to (or towards) this maximum gives the set of parameter values to use for the next iteration. It requires the calculation of the Hessian—the matrix of second-order partial derivatives of the log-likelihood, with respect to the model parameters—and this can be a computationally expensive procedure. A *quasi-Newton method* avoids this complication by instead using successive first-order partial derivatives to approximate the Hessian. Not only is this far simpler to compute, but its use can—paradoxically—result in a better-performing algorithm than using the true Hessian; see Press, Teukolsky, Vetterling, and Flannery (2002, pp. 521–524) for further information.

While quasi-Newton algorithm implementations exist in common R functions (e.g., the `"BFGS"` method of the `optim()` function from the `stats` package), the advantage[1] that ADMB holds over these is that it computes the first-order derivatives analytically via *automatic differentiation* (AD), rather than via a finite-difference approximation. AD is achieved by applying the chain rule for differentiation successively for each binary operation and mathematical function specified in the calculation of the objective function (see Griewank & Walther, 2002). Its use has two substantial advantages, being (i) analytic differentiation is not subject to the error associated with a numeric approximation, resulting in more accurate derivatives and therefore more precise optimisation with higher efficiency; and (ii) the approximation of each partial derivative by finite differencing requires an additional function evaluation and so fitting models with larger numbers of parameters can incur substantially larger computational costs. Using AD, on the other hand, only requires a single function evaluation regardless of the number of model parameters (although this comes with the additional overhead of the gradient calculations). One would therefore expect an optimiser that implements AD to perform more efficiently, especially when a large number of parameters are to be estimated.

Upon execution of the `admbsecr()` command, the data provided as the `capt` argument, along with a variety of other model options (e.g., the types of auxiliary data held within

[1] An additional advantage is that a compiled language (such as C++) will generally benefit from faster performance over an interpreted language (such as R). However, this can be mitigated via calculation of the likelihood using an R function that calls compiled C++ code. The `Rcpp` package is an excellent tool that facilitates the integration of R and C++ in this way.

`capt`), are written to the `.dat` file. Start values for each parameter are written to the `.pin` file. These are generated automatically, but the defaults can be overridden by specifying the optional `sv` argument of `admbsecr`. This takes a list, where each component name is that of a parameter, and the component itself is a scalar providing the chosen start value. See the parameter entries in the Notation list at the end of this thesis for the names used by `admbsecr`.

A variety of other optional arguments are available to the user that can potentially improve optimisation convergence or efficiency. For example, the `bounds` argument can be specified as another list, where each component again corresponds to a parameter. Each consists of a vector of length two, specifying a lower and upper bound for the parameter in question; this allows the user to impose a restricted parameter space. Additionally, parameters can be introduced to the optimisation process in phases. During the first phase only a subset of parameters are maximised over, leaving those remaining at their start values. Remaining parameters can then be introduced, one phase at a time. These phases can be specified in the argument `phases` as a list. The `fix` argument can be used to fix parameters at particular values, without ever including them for maximisation. These arguments are passed to the ADMB template via the `.dat` file.

The ADMB template used to generate the executables used for optimisation can be found in the file `ADMB/src/secr.tpl`, with this file specification being relative to the main directory of the installed package. Precompiled executables for Windows, OSX, and Linux platforms can be found within the `ADMB/bin/` directory. The `admbsecr()` function determines the operating system in use and calls the appropriate executable for optimisation.

Likelihood computation

The n integrals over detected animals' locations (Equation (2.3)) are not available in closed form, and therefore must be approximated numerically. Following the recommendation of Borchers and Efford (2008), this is done via the so-called rectangle rule, whereby the domain $\mathcal{A}$ is divided into a grid of small squares. The grid of square centre points is often known as the *habitat mask*. The value of the integrand is computed at each mask point; this value is used as an approximation over its entire square. The area of the square can then be multiplied by this value to give its contribution to the integral, and this is summed across all mask-points—that is, using a as the area of each square and $\mathcal{M}$ to represent the set of

all mask points,

$$\prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}_i \approx$$

$$\prod_{i=1}^{n} a \sum_{\boldsymbol{x}_i \in \mathcal{M}} f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma}). \quad (2.15)$$

Integration over the entire survey region, $\mathcal{A}$, may not necessarily be required: some regions of $\mathcal{A}$ may be so far from the detectors such that any individual with an associated location therein is virtually undetectable. Thus, it is known *a priori* that $f(\boldsymbol{x}) \approx 0$ if $\boldsymbol{x}$ lies within this region, and so there is no need to include it in the domain over which integration occurs. One way to achieve this is by specifying some *buffer distance*—a distance beyond which a single detector is not feasibly able to detect any individuals. The mask points can then be arranged so that each is within distance b of at least one detector. Reducing b too far will cause negative error in the above approximation, while setting it too large will inflate computation time. The above approximation converges to the correct value as the resolution of the mask increases—or, equivalently, as $a \to 0$, and $b \to \infty$. No mask points should lie within areas that are known to be uninhabited (e.g., due to an unsuitable habitat type). All existing ML SECR software implementations use the same approximation shown above.

The habitat mask used to fit a given model can be specified using the `mask` argument of the `admbsecr()` function. It constitutes a matrix, where each row gives the coordinates of a mask point. The function `create.mask()` facilities its construction, taking the detector locations and the buffer as arguments, and returning a suitable object. Figure 2.1 shows the detector and mask point locations used for analysis of the `example` data.

One naïve implementation of code that computes (the log of) Equation (2.15) would involve a loop that iterates over each individual. Each iteration would approximate the integral for each, using a nested loop to calculate the three terms in the integrand for each mask point, and keeping record of the cumulative sum. This simple approach can be improved upon in three ways to increase computational efficiency—in some cases substantially so.

First, for any two individuals, the final term in the integrand is equivalent upon consideration of the same mask point. Therefore, $f(\boldsymbol{x}; \boldsymbol{\gamma})$ need only be calculated once for each element of $\mathcal{M}$, then saved and reused for each individual.

Second, the second term in the integrand is equivalent across all mask points for any two individuals with the same capture history. Thus, it is sensible to implement a loop

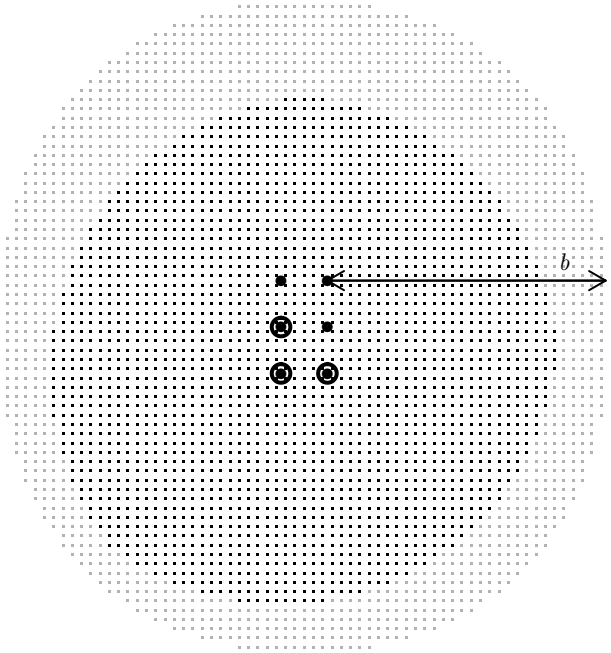


Figure 2.1 *Detector (large points) and mask point (small points) locations used for the* `example` *data.The mask was generated by* `create.mask()`*, setting the buffer distance, b. Dark mask points are those that would be used with local integration for individuals detected by the circled detectors. Light mask points are those that would be ignored.*

that iterates over each *unique* capture history. At each iteration the term $f(\boldsymbol{\omega}_i|\boldsymbol{x}_i;\boldsymbol{\gamma})$ is calculated for every mask point. A nested loop then iterates over each individual with this particular capture history, within which only $f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i,\boldsymbol{x};\boldsymbol{\psi})$ requires calculation for every mask point. The precalculated $f(\boldsymbol{\omega}_i|\boldsymbol{x}_i;\boldsymbol{\gamma})$ is then combined with this term to calculate the integrand at each mask point—and therefore approximate the integral. This prevents needless recalculation of $f(\boldsymbol{\omega}_i|\boldsymbol{x}_i;\boldsymbol{\gamma})$ for individuals with the same capture histories.

This approach is particularly effective at reducing processing time if the number of detectors is small. The maximum number of unique capture histories is given by $2^m - 1$. For example, if $m = 3$, there are (at most) seven different values of $f(\boldsymbol{\omega}_i|\boldsymbol{x}_i;\boldsymbol{\gamma})$ that require calculation at each mask point. Naïve computation by looping over individuals results in n such calculations.

Finally, the number of mask points at which calculation of the integrand takes place for each individual can be considerably reduced using *local integration*, unique to the `admbsecr` implementation of SECR software. Recall that the buffer distance, b, should be set so that an animal further than this distance should be undetectable. Thus, if this holds, there is no

need to integrate over any regions of the spatial domain that are not within distance b of all detectors that made a detection; that is, rather than integration over $\mathcal{A}$ for all detected individuals, integration can be restricted to a region $\mathcal{A}_i \subset \mathcal{A}$ for the ith individual, where

$$\mathcal{A}_i = \bigcap_{\{j:\omega_{ij}=1\}} \{\boldsymbol{x} : d_j(\boldsymbol{x}) < b\}\,.$$

See Figure 2.1 for an example of the effect of local integration on the `example` data.

Local integration has the greatest effect for surveys with large arrays of detectors, and when their spread across the survey region is far greater that the range of detectability (and therefore far greater than b)—in this case, each $\mathcal{A}_i$ is only a small subset of $\mathcal{A}$, leading to a substantial reduction in computing cost. However, for small detector arrays, the computational overhead concerned with determining each $\mathcal{A}_i$ can be greater than the savings in processing time. Furthermore, its use greatly heightens the sensitivity of $\widehat{\boldsymbol{\theta}}$ to values of b that are too small: if some individuals are detected from distances greater than b, then some domains of integration are potentially far too restrictive; this leads to inaccurate approximation of the corresponding integrals. For these reasons local integration is optional, and can be set via the logical `admbsecr()` argument `local`.

The `admbsecr()` function

The `example` object is a list consisting of a number of components, including those named `capt`, `traps`, and `mask`. These are appropriate for use as the corresponding arguments to the `admbsecr()` function. In particular, `capt` is a list itself, containing all kinds of previously mentioned auxiliary information. The `example` list has an additional component, `cutoff`, that is appropriate for use as the threshold value, c.

The following code can be used to fit a model to these data, incorporating all auxiliary information types:

```
fit <- admbsecr(capt = example$capt, traps = example$traps,
                mask = example$mask,
                ss.opts = list(cutoff = example$cutoff), trace = TRUE)
```

Setting the `trace` argument to `TRUE` is useful if fitting a model interactively; it results in the parameter values at each step of the optimisation procedure being printed to the screen, along with occasional updates of the partial derivatives of the log-likelihood.

The resulting model object, `fit`, is itself a list. It contains various information about the fitted model. Rather than the user inspecting this themselves, the best way to access

the information held within is to use a variety of utility and plotting functions that are exported by the `admbsecr` package. These are explained in further detail throughout the rest of this chapter—primarily Sections 2.3.2 and 2.3.4.

2.3.2 Utility functions

The `admbsecr` package exports various S3 methods for generic functions that can be used to extract model information.

The `coef()`, `stdEr()`, and `summary()` functions provide parameter estimates, standard errors, and a model summary, respectively. The latter provides estimates and standard errors, along with additional model information. It also returns the ESA, $a(\boldsymbol{\gamma})$ (Equation (1.6)), as a derived parameter, along with its standard error:

```
summary(fit)

# Detection function: Signal strength
# Information types: Bearings, Distances, Signal strengths, Times of arrival
#
#  Parameters:
#             Estimate Std. Error
# D         2.3993e+03   227.7800
# b0.ss     9.0016e+01     1.3914
# b1.ss     3.9698e+00     0.2045
# sigma.ss  9.7630e+00     0.4826
# kappa     4.8294e+01     6.0555
# alpha     4.9416e+00     0.4601
# sigma.toa 1.8531e-03     0.0001
# ---
# esa       5.2931e-02     0.0018
```

The variance-covariance matrix and CIs for the parameters can be obtained using `vcov()` and `confint()`, respectively:

```
confint(fit)

#                  2.5 %       97.5 %
# D         1.952885e+03 2.845767e+03
# b0.ss     8.728940e+01 9.274359e+01
```

```
# b1.ss     3.568966e+00 4.370630e+00
# sigma.ss  8.817119e+00 1.070880e+01
# kappa     3.642536e+01 6.016249e+01
# alpha     4.039744e+00 5.843421e+00
# sigma.toa 1.610651e-03 2.095546e-03
```

The former is not shown here for brevity. The default confidence level is 95%, but this can be adjusted via the optional `level` argument.

Finally, the log-likelihood at the estimated parameter values and the Akaike information criterion (AIC) can be extracted using `logLik()` and `AIC()`:

```
logLik(fit)

# 'log Lik.' -1281.74 (df=7)

AIC(fit)

# [1] 2577.48
```

2.3.3 Variance estimation

ADMB estimates the Hessian following optimisation. Default behaviour is for calculation of the variance-covariance matrix (using the `vcov()` function) via the inversion of the Hessian[2]. Standard errors (returned by `stdEr()` and `summary()`) are obtained by taking the square-root of the main diagonal of this matrix. Wald CIs, relying on normal approximations of the parameters' sampling distributions, are then returned by `confint()`.

In some cases this normal approximation may not be appropriate (e.g., if the number of detections is not yet large enough for the asymptotic normality of parameter sampling distributions to apply). In such cases, variance estimation via a parametric bootstrap procedure may be considered more appropriate. The `admbsecr` package provides implementation of such a procedure in the `boot.admbsecr()` function.

The main argument of this function is a fitted model object (e.g., `fit`, created above), and the number of bootstrap resamples can be selected using the argument `N`. Naturally, this can be a computer-intensive process, especially when `N` is large. This function therefore supports parallel processing: the argument `n.cores` allows the user to select the number

[2]This is not inversion of the observed information matrix—the *negative* of the estimated Hessian—as ADMB does not maximise the log-likelihood; rather, it minimises the negative log-likelihood.

of processor cores on which to simultaneously fit the model in question to resampled data. Example code for execution of the bootstrap procedure is therefore

```
boot.fit <- boot.admbsecr(fit = fit, N = 1000, n.cores = 4)
```

Following this procedure, the variance-covariance matrix, all standard errors, and CIs are calculated from the parameter estimates based on the resampled data. Observe how the following functions return different (albeit similar) values when the `boot.fit` is used as the main argument:

```
summary(boot.fit)

# Detection function: Signal strength
# Information types: Bearings, Distances, Signal strengths, Times of arrival
#
#  Parameters:
#             Estimate Std. Error
# D         2.3993e+03   221.0511
# b0.ss     9.0016e+01     1.3189
# b1.ss     3.9698e+00     0.1945
# sigma.ss  9.7630e+00     0.4716
# kappa     4.8294e+01     5.2182
# alpha     4.9416e+00     0.4318
# sigma.toa 1.8531e-03     0.0001
# ---
# esa       5.2931e-02     0.0018

confint(boot.fit)

#                   2.5 %       97.5 %
# D         1.966074e+03 2.832578e+03
# b0.ss     8.743151e+01 9.260148e+01
# b1.ss     3.588653e+00 4.350943e+00
# sigma.ss  8.838586e+00 1.068733e+01
# kappa     3.806652e+01 5.852133e+01
# alpha     4.095260e+00 5.787906e+00
# sigma.toa 1.610602e-03 2.095595e-03
```

Standard errors for each parameter are calculated from the standard deviations of the bootstrap resample parameter estimates. A choice of bootstrap-based CI methods—all described by Davison and Hinkley (see 1997)—are available, and can be chosen using the `method` argument.

The default selection is to simply replace the standard errors calculated from the Hessian with those from the bootstrap procedure in the calculation of the Wald CIs. Bootstrap procedures are capable of estimating an estimator's bias by subtracting the estimate from the observed data from the mean of the estimates from the bootstrap resamples. Estimated bias for each parameter can be extracted using the `get.bias()` function, using the bootstrapped fitted model object as its argument. A bias-corrected version of the default interval is available by setting the `method` argument to `"default.bc"`. In order to check the appropriateness of the normal approximations, a quantile-quantile (QQ) plot for each parameter can be generated by setting the argument `qqplot` to `TRUE`. So-called percentile and basic CIs can be obtained by setting `method` to `"percentile"` and `"basic"`, respectively.

The nature of simulation-based methods like the bootstrap is such that a repetition of the procedure will not provide equivalent results; the results are subject to the generation of (pseudo)random values, and the simulation process necessarily has a finite number of iterations. It is therefore important to quantify the *Monte Carlo error* (MCE) associated with any estimates generated from such an approach. This is defined as the standard deviation of the estimator over hypothetical repetition of the simulation process; this naturally decreases as the simulation size increases.

The MCE associated with standard errors and biases can be evaluated with use of a secondary bootstrap (Koehler, Brown, & Haneuse, 2009). This initially may seem computationally infeasible; however, it only requires taking nonparametric bootstrap resamples (i.e., resampling with replacement) from the *already-obtained* bootstrap resample parameter estimates. Thus, no further models need be fitted, and so this constitutes only a small portion of the original bootstrap procedure's processing time. Estimated MCEs for each parameter are given by the standard errors across this second stage of resampling (see Equation (9) in Koehler et al., 2009). The `boot.admbsecr()` function runs this second bootstrap procedure upon completion of the first. Its argument `M`, which defaults to 10 000, sets its number of resamples. Extracting MCEs for both standard errors and estimated biases can be achieved by setting the `mce` argument to `TRUE` upon calling the `stdEr()` and `get.bias()` functions:

```
stdEr(boot.fit, mce = TRUE)

#             Std. Error          MCE
# D         2.210511e+02 5.102717e+00
# b0.ss     1.318895e+00 3.146790e-02
# b1.ss     1.944653e-01 4.576561e-03
# sigma.ss  4.716275e-01 1.098582e-02
# kappa     5.218160e+00 1.274679e-01
# alpha     4.318054e-01 1.029445e-02
# sigma.toa 1.237251e-04 2.786864e-06

get.bias(boot.fit, mce = TRUE)

#                    Bias          MCE
# D         -2.433555e+00 6.994455e+00
# b0.ss      2.507345e-01 4.218931e-02
# b1.ss      2.799561e-02 6.193708e-03
# sigma.ss  -1.223925e-02 1.506966e-02
# kappa     -7.743596e+00 1.647840e-01
# alpha     -1.510141e-01 1.367849e-02
# sigma.toa  3.404924e-05 3.889485e-06
```

In this case, for example, the MCEs for all parameters are just over 2% of their respective standard errors. If this is not sufficient, the user may elect to repeat the initial bootstrap procedure with a larger number of resamples to decrease the MCEs further.

2.3.4 Plotting functions

Additional functions can aid the understanding of the model output by creating plots. The `show.survey()` function displays the detector and mask locations; see Figure 2.1. The `show.detfn()` function takes the fitted model object as its main argument and plots the estimated detection function; see Figure 2.2. The `show.detsurf()` function plots the DPS, $p(\boldsymbol{x};\boldsymbol{\gamma})$. The `locations()` function plots estimated locations of individuals.

This the latter is via derivation of the estimated density of a location $\boldsymbol{x}_i$, given its observed capture history and auxiliary information. Borchers et al. (2015) showed that this

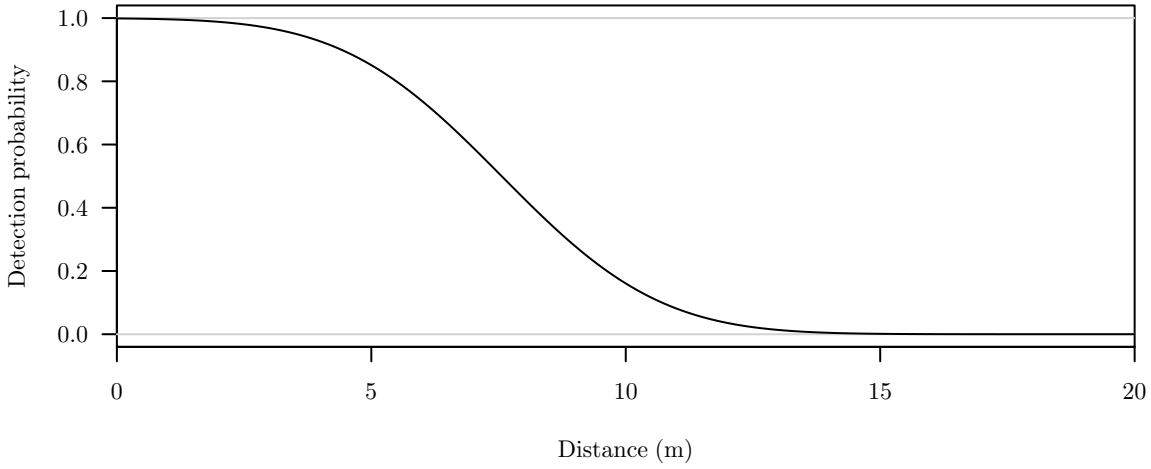


Figure 2.2 *The estimated detection function from the* `example` *data, generated using the* `show.detfn()` *function. It is estimated that individuals within 2 m of a detector are almost certainly detected, whereas those situated further than 15 m are virtually undetectable.*

can be achieved with an application of Bayes' theorem:

$$f(\boldsymbol{x}_i|\boldsymbol{\omega}_i, \boldsymbol{Y}_i; \boldsymbol{\gamma}, \boldsymbol{\psi}) = \frac{f(\boldsymbol{\omega}_i, \boldsymbol{Y}_i|\boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\psi})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})}{\int_{\mathcal{A}} f(\boldsymbol{\omega}_i, \boldsymbol{Y}_i|\boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\psi})\, f(\boldsymbol{x}; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}}$$
$$= \frac{f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})}{\int_{\mathcal{A}} f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}; \boldsymbol{\gamma})\, f(\boldsymbol{x}; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}}.$$

The numerator (and integrand of the denominator) are, in fact, simply the integrand of the integral in the likelihood that is evaluated for each individual; see Equation (2.3). This PDF can be plotted over the spatial domain $\mathcal{A}$ to provide an estimate of the location of the ith individual along with its associated uncertainty (although it is important to note that this is conditional on the parameter values, and therefore does not incorporate any parameter uncertainty).

Aside from being of inherent interest, it is also useful for evaluating the improved precision in location estimates due to the use of auxiliary information. For example, Figure 2.3 shows representations of PDFs that both do and do not incorporate auxiliary information. In this case—due to the wealth of data collected about each detection—the location estimate is very precise when the auxiliary information is incorporated.

2.3.5 Data simulation

Simulation of SECR detection data (with or without any type of auxiliary information mentioned above) can be achieved using the `sim.capt()` function. This can either be done

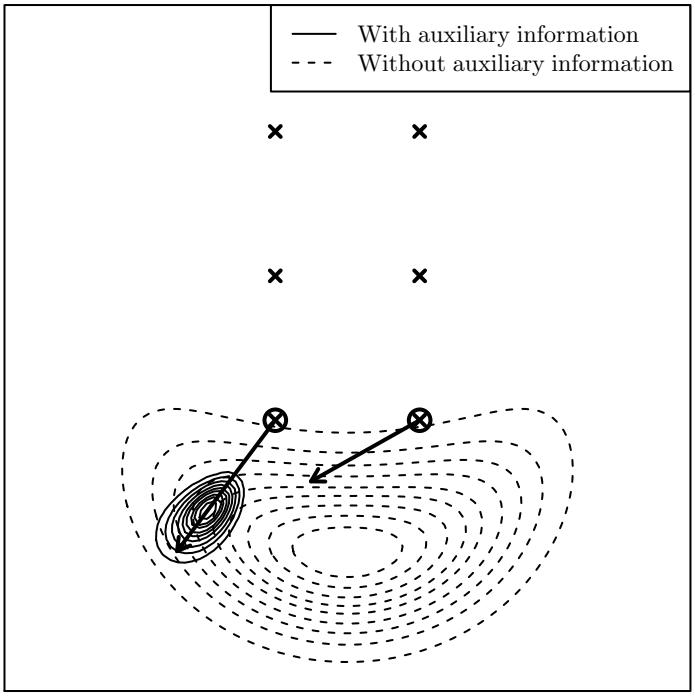


Figure 2.3 *Estimated conditional PDFs of the first detected individual's location, both with and without incorporation of auxiliary information. Crosses represent detector locations, while those that are circled detected this particular individual. Arrows show both the estimated bearings (through their directions) and estimated distances (through their lengths) from each detector. Signal strength and TOA information is not explicitly represented, but is nevertheless used in the calculation of the solid contours. It is estimated that the individual is located within the outermost contour of each set with a probability of 0.9, and is located within the innermost contour of each set with a probability of 0.1.*

by simply providing a fitted model object (in which case data are simulated from the same model at the estimated parameter values), or by specifying the auxiliary information types required along with all necessary parameters.

For the latter, many of the arguments are equivalent to those used by `admbsecr()`; this includes `traps`, `mask`, `detfn`, and `ss.opts`. The `infotypes` argument must be a vector of character strings, each indicating a auxiliary information type to simulate (i.e., a combination of `"dist"`, `"bearing"`, `"ss"`, and `"toa"`). The argument `pars` must be a named list, where parameter names correspond to parameters, and the components themselves provide parameter values.

The `sim.capt()` function begins by simulating animal locations as a realisation of a Poisson point process with homogeneous intensity D. For each individual, detections across the m detectors are then generated from the untruncated PMF $f(\boldsymbol{\omega}_i|\boldsymbol{x}_i;\boldsymbol{\gamma})$ (i.e., including the probability of nondetection). Individuals that are not detected by any detectors are

discarded. Each auxiliary information type is then simulated from $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$ for all detectors that detected the individual in question. All detection data are then returned in a format suitable for use as the `capt` argument of the `admbsecr()` function. If the user wishes to keep the simulated animal locations the argument `keep.locs` can be set to `TRUE`.

2.3.6 Compatibility with `secr`

The `admbsecr()` package is designed to minimise overlap with the widely used `secr` package, with the vision that the two can be used side-by-side by the same user with minimal effort expended familiarising themselves with two R packages. In the interests of making `admbsecr` more readily usable for existing `secr` users, a number of functions are available that convert objects suitable for use with functions from the `secr` package to those for use with functions from the `admbsecr` package. The function `convert.capt.to.admbsecr()` will take the analogous capture-history argument of the `secr.fit()` function (used to fit SECR models with the `secr` package) and return an object suitable for the `capt` argument of `admbsecr()`. The `convert.capt.to.secr()` function will perform this conversion in the opposite direction. One note of warning is that conversion should not be attempted for objects containing any type of auxiliary information other than received signal strengths. This is because the `secr` package is not compatible with these other information types.

The most straightforward way of creating objects containing detection information from scratch is via the so-called `captures` data frame, and this holds for both the `admbsecr` and the `secr` packages. Each row corresponds to the detection of a single animal by a single detector. The second and fourth columns should provide the identification of the animal and the detector corresponding to the detection, respectively. The fifth column onwards can be used for auxiliary information, with column names of `"dist"`, `"bearing"`, `"ss"`, and `"toa"` corresponding to the appropriate information type. This data frame can then be provided as an argument to the `make.capthist()` (from the `secr` package) or `create.capt()` (from the `admbsecr` package), giving rise to capture objects suitable for use with the `secr.fit()` and `admbsecr()` functions, respectively. The first and third columns in the `captures` data frame correspond to temporal capture information, and are thus used by `make.capthist()` but ignored by `create.capt()`. Further details are available in the data structure vignette referred to earlier.

2.3.7 Unit testing

In order to aid the continued development and upkeep of the `admbsecr` package, unit testing has been implemented via the `testthat` package. Although not typically run by users, these

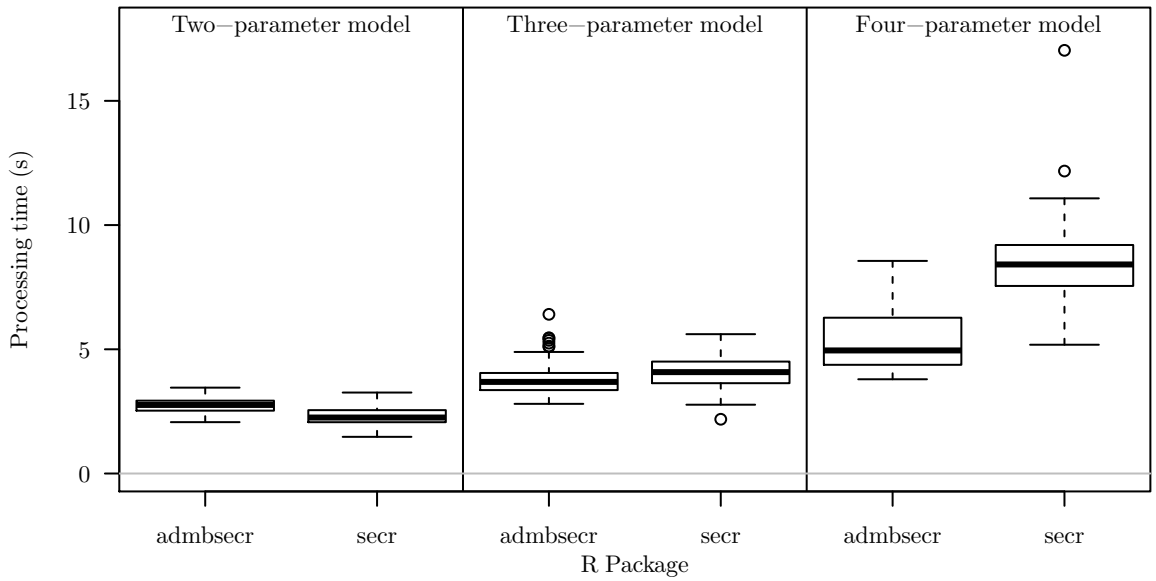


Figure 2.4 *Boxplots showing parameter estimation processing times using* `admbsecr` *and* `secr` *for the three benchmarked models. The R package* `admbsecr` *was faster on average for both the three- and four-parameter models, incurring a smaller cost with each additional parameter estimated.*

tests allow for straightforward identification of bugs during software development. Tests can be run using `test.admbsecr()`. Setting the sole argument `quick` to `TRUE` will not run the full suite of tests, but will check compatibility of the ADMB executable on the user's system.

2.3.8 Benchmarking

In order to evaluate the performance of ADMB as an optimiser, the `admbsecr` package was benchmarked against the `secr` package across three models: (i) SECR using the half-normal detection function (with the g_0 parameter fixed at 1), a two-parameter model; (ii) SECR using the hazard-rate detection function (with the g_0 parameter fixed at 1), a three-parameter model; and (iii) SECR with auxiliary signal strength data, a four-parameter model. Models incorporating further auxiliary data could not be used for this comparison, as these do not fall within the scope of the `secr` package.

For each model, 100 sets of detection data were generated by the `sim.capt()` function. Both the `secr` and `admbsecr` package were used to fit the same model to each, and the processing time taken to do so was recorded. The `admbsecr` package outperformed the `secr` package for all but the two-parameter model (Figure 2.4); its relative performance increased with the number of model parameters. This is unsurprising—optimisation that benefits from AD should expect to perform relatively better with increasing model complexity.

2.4 Applications

The use of the model class developed in this chapter is demonstrated through its application to detection data from surveys of three different species: *B. acutorostrata*, *N. annamensis*, and *A. lightfooti*. These surveys are described below, and are also accompanied by descriptions of the analyses and a relevant simulation study, along with the corresponding results.

2.4.1 Common minke whale

Two independent observers aboard a plane surveyed the same area of ocean and recorded sightings of *B. acutorostrata* surfacings. For each detection, the perpendicular distance from the plane to the sighted whale was estimated using a declinometer. These data were collected during the 2001 North Atlantic Sightings Survey; see Pike, Paxton, Gunnlaugsson, and Vikingsson (2009) for further details.

Data like these are typically analysed using an MRDS model, and in this case distances must be assumed to be exact. An issue here is that when the observers sighted the same animal, their estimated distances routinely differed to some degree. This is typically reconciled by simply averaging the differences and assuming this mean difference is the exact, true difference. Here, an SECR model is fitted to these data using the auxiliary distance information. This allows for the observed distances to be modelled, and estimation of their associated measurement error.

In total, 70 whales were detected by at least one of the two observers. There was strong evidence to suggest that one observer was far more proficient in whale detection, and so a separate detection function was estimated for each[3]. An SECR model—estimating distance estimation error—provided a whale density estimate of 1.72 whales per hectare. An MRDS model—taking the distance estimates to be error free—provided a whale density estimate of 1.61 whales per hectare. Both estimates were associated with similar standard errors.

One particularly interesting feature of the SECR analysis is that the detectors (i.e., the observers) are considered to have the same location, and—for traditional SECR models—the detectors must be spatially discrete in order to estimate the detection function. Here it is shown that this is not a requirement for SECR, so long as spatial information is available from another source—in this case the estimated distances.

In order to ascertain the effect of modelling the error in the estimation of distances, a

[3]This analysis was carried out in an outdated version of `admbsecr`. At the time of writing, `admbsecr` no longer supports detector-specific detection function estimation. This may be reimplemented in the near future.

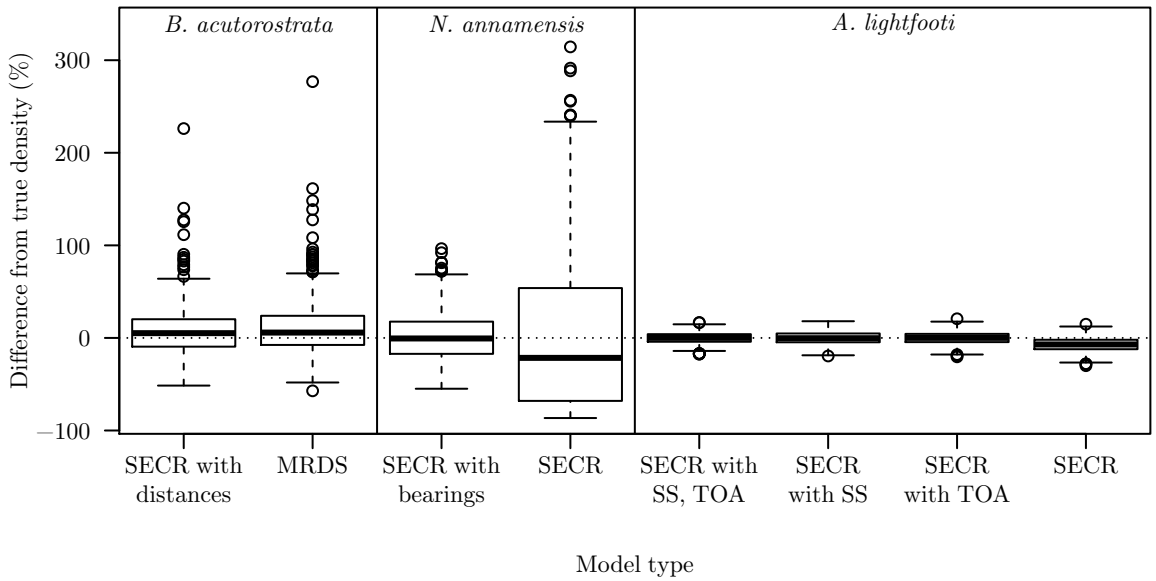


Figure 2.5 *Boxplots showing estimated* B. acutorostrata*,* N. annamensis*, and* A. lightfooti *densities from the simulated data sets for (i) the SECR and MRDS models (for* B. acutorostrata*), and (ii) the SECR models that do and do not incorporate auxiliary observed data (for* N. annamensis *and* A. lightfooti*). Estimates are shown as percentage differences from the underlying density used to generate the data.*

simulation study was carried out. In total, 500 data sets were simulated, using parameter values similar to those estimated from the real data. Each was analysed using both an SECR model and an MRDS model (averaging the estimated distances for individuals detected by both observers). Both SECR and MRDS whale density estimators showed slight positive bias; however, the former had lower variance, and therefore also a lower mean-squared error (MSE; Figure 2.5, Table 2.1). The bias is suspected to be small-sample bias, as six parameters (whale density, distance measurement error, and two detection-function parameters for each observer) were estimated from detections of only approximately 70 animals, on average.

2.4.2 Northern yellow-cheeked gibbon

In 2010, an acoustic survey of *N. annamensis* was conducted by Conservation International in the jungles of northeastern Cambodia. Three observers were stationed in a line, 500 m apart, and recorded detections of groups of calling gibbons. For each detection, the observer in question recorded an estimated bearing to the gibbon group. See Kidney et al. (in submission) for further details.

Here, these estimated bearings are incorporated into an SECR model as auxiliary in-

Table 2.1 *Estimated biases, variances, and root mean square errors from the whale and gibbon simulation studies. All are given as percentages of the true underlying animal density used to generate the data.*

Species	Model type	Bias (%)	SD (%)	RMSE (%)
B. acutorostrata	SECR with distances	8.23	27.07	28.29
	MRDS	10.79	30.80	32.63
N. annamensis	SECR with bearings	1.56	25.88	25.93
	SECR	4.29	86.76	86.86
A. lightfooti	SECR with SS, TOA	0.14	6.15	6.16
	SECR with SS	0.15	6.76	6.76
	SECR with TOA	−0.02	6.75	6.75
	SECR	−6.93	7.39	10.13

formation, resulting in an estimated gibbon group density of 0.319 groups km^{-2} (with a standard error of 0.074). A model that ignores the estimated bearings generates an estimate of 0.8290 groups km^{-2} (with a substantially larger standard error of 0.367).

A simulation study was carried out to investigate the effect of the incorporation of estimated bearings on the animal density estimator. As before, 500 data sets were simulated, with parameter values set at values similar to those estimated from the real data. Introducing the use of estimated bearings considerably reduces estimator variance (Figure 2.5, Table 2.1), and this due to the additional spatial information they provide (Figure 2.6). See Kidney et al. (in submission) for a rigorous analysis of these data, and further related methodological developments particular to the estimation of gibbon group density via SECR.

2.4.3 Cape Peninsula moss frog

See Section 3.5.1 for a detailed description of the *A. lightfooti* survey data. Briefly, six microphones were placed within a montane seepage at various sites, and male advertisement calls were recorded. The microphones were connected to a central audio recorder, allowing for the collection of relative TOAs of each detected call across the microphone array. Signal strengths were also observed for each detection. See Chapter 3 for further details.

Developing further methodology for the analysis of data from these surveys provides motivation for large portions of Chapter 3, where a rigorous treatment of the analysis of data from such a survey can also be found. Results from models fitted to these observed data are therefore not shown here, but this particular application is nevertheless mentioned

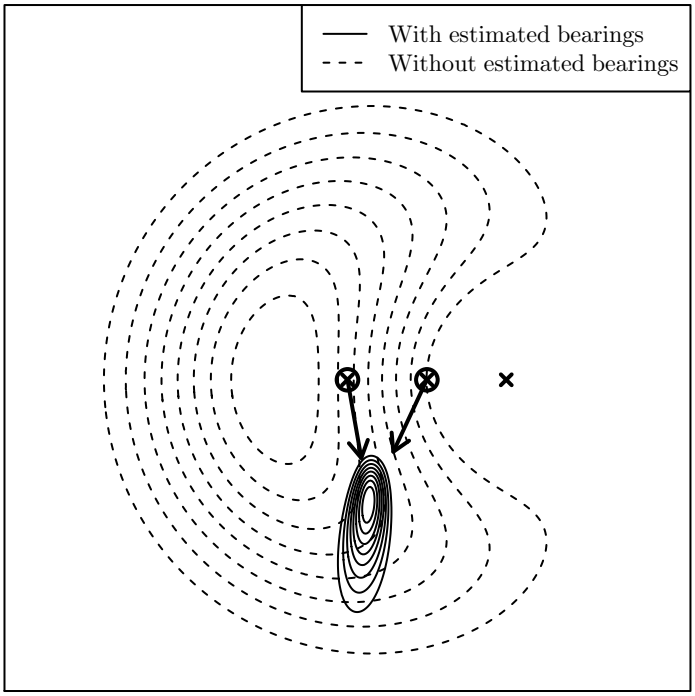


Figure 2.6 *Estimated conditional PDFs of a detected group's location—both with and without incorporation of its estimated bearings. Arrows show the estimated bearings from from observers that detected the individual. It is estimated that the individual is located within the outermost contour of each set with a probability of 0.8, and is located within the innermost contour of each set with a probability of 0.1.*

in order to further highlight the flexibility of the model class discussed.

Once more, a simulation study was carried out to investigate the effect of the use of the auxiliary information—in this case, the signal strength and TOA data. Incorporation of these data again was shown to reduce estimator variance (Figure 2.5, Table 2.1).

The models that did not incorporate the signal strength data fitted a half-normal detection function; however, the data were generated using the signal strength detection function. This therefore constitutes a model misspecification, and likely is the reason for the observed bias in the SECR model that did not use any auxiliary data. It is interesting that this bias was virtually eliminated following the incorporation of TOA data.

A separate simulation study (not shown here for brevity) indicated that this bias was also eliminated by the use of the threshold detection function when neither signal strength nor TOA information were available.

2.5 Discussion

This chapter has demonstrated that the incorporation of auxiliary spatial information can substantially improve the precision of SECR animal density estimators (Figure 2.5, Table 2.1). This improvement is greatest when the detector locations alone are not particularly informative about detected individuals' locations. The survey of *N. annamensis*, for example, only deployed three observers, and so there were only $2^3 - 1 = 7$ possible observed capture histories. Knowledge of which detectors detected a particular gibbon group is relatively uninformative—but the measurement error associated with the auxiliary information is sufficiently small such that its incorporation results in fairly precise location, and therefore density, estimates (Figure 2.6).

Additionally, there is a cost incurred when assuming that auxiliary data are more precise than they actually are: modelling the measurement error in distances to *B. acutorostrata*—rather than assuming they were exact—resulted in an improved animal density estimator (Figure 2.5, Table 2.1). Although the improvement was modest, variance estimates are probably greatly affected by this misspecified assumption. Failing to recognise a source of variation in the data likely causes negative bias in standard errors, and therefore CIs that do not have sufficient coverage. This effect, however, is yet to be formally investigated.

2.5.1 Further model development

As presented in Section 2.2.1, estimated distances are assumed to be unbiased. There is justification for this in the application to *B. acutorostrata*: the distances were measured with use of a declinometer, and so it may be reasonable to expect this to give the correct distance, on average. In other contexts this assumption may not be reasonable.

For example, there is potential for observers to estimate distances in addition to bearings on the gibbon survey. In this case, however, it is simply a subjective distance estimate with no calibration whatsoever from the use of any kind of tool. Assuming unbiasedness may not be reasonable. Despite this, distance estimates remain informative; an individual is likely to be closest to an observer with the smallest distance estimate, and furthest from the one with the largest distance estimate—even if these estimates are biased. As spatial information also comes from other sources (the detector locations, and potentially estimated bearings), it is possible to determine the existence of any systematic over- or underestimation of distances. Estimation of this bias could be achieved by setting Equation (2.5) to

$$\mathrm{E}(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi},\tau) = d_j(\boldsymbol{x}_i) - \tau, \text{ or}$$

$$\mathrm{E}(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi},\tau) = \frac{d_j(\boldsymbol{x}_i)}{\tau},$$

where τ represents either absolute or proportional bias, respectively. For the former, this gives

$$f(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi}) = \frac{\alpha^\alpha\, y_{ijk}^{\alpha-1}\, \exp(-\alpha\, y_{ijk}/[d_j(\boldsymbol{x}_i)-\tau])}{[d_j(\boldsymbol{x})-\tau]^\alpha\, \Gamma(\alpha)}.$$

Alternative extensions involve providing discrete (i.e., 'binned') distance estimates (e.g., '$<$ 100 m', '100–200 m', etc.), thus resulting in a discrete distribution for the conditional random variable $y_{ijk}|\boldsymbol{x}_i$. The PMF of a binned distance estimate (incorporating a bias correction) then becomes

$$f(y_{ijk}|\boldsymbol{x}_i;\boldsymbol{\psi}) = \int_{l(y_{ijk})}^{u(y_{ijk})} \frac{\alpha^\alpha\, y^{\alpha-1}\, \exp(-\alpha\, y/[d_j(\boldsymbol{x}_i)-\tau])}{[d_j(\boldsymbol{x})-\tau]^\alpha\, \Gamma(\alpha)}\, \mathrm{d}y,$$

where $u(y_{ijk})$ and $l(y_{ijk})$ provide the upper and lower limits of the binned distance estimates, respectively.

Equivalent procedures are theoretically possible for estimated bearings. However, an assumption of unbiasedness is far more reasonable here due to symmetry.

While only four specific types of auxiliary data were discussed here, the framework for incorporation of further types has been presented—all that is required is the derivation of $f(\boldsymbol{y}_{i\cdot k}|\boldsymbol{\omega}_i,\boldsymbol{x}_i;\boldsymbol{\psi})$. There is potential for an interesting extension of this model class to surveys that employ physical capture techniques; for example, a trapped animal being wet may suggest that its burrow is located across a river, a seed trapped in its fur may indicate that it has recently travelled through a particular area densely populated by this plant species, or inspection of its scat may provide information about its feeding grounds. The derivation of the appropriate likelihoods for each situation is likely to vary according to the survey region, the focal species, and the type of traps deployed.

2.5.2 Extensions to `admbsecr`

The use of ADMB was initially motivated by its powerful capability of modelling latent variables using the Laplace approximation. This provides an extremely efficient means of approximating their marginalisation out of the joint density with the observed data (e.g., in the case of SECR, the integrals in Equations (1.12) and (2.3)). The Laplace approximation operates by finding the mode of the integrand (e.g., the functions represented by the contours in Figures 2.3 and 2.6) with respect to the latent variables, before using a Gaussian function centred on this point as an approximation. Gaussian integrals are

available in closed form and are thus trivial to compute. However, an implementation of the Laplace approximation to the calculation of the SECR likelihood presented in this chapter routinely resulted in optimiser nonconvergence. In cases where convergence was obtained, the resulting estimates were not always comparable to those found using the standard quadrature integration method.

Naturally, the accuracy of the Laplace approximation suffers the further the integrand departs from a Gaussian function. The associated error is unacceptably large with marked departures—for example, in cases where the integrand is severely skewed or multimodal. The latter is particularly prevalent for integration over animals' locations in SECR: consider an individual detected only at opposite ends of a detector array; it is likely to be close to either end, but is unlikely to be located in between due to the presence of detectors that failed to make a detection. The use of the Laplace approximation for the calculation of SECR likelihoods therefore appears unsuitable—a Gaussian function does not sufficiently approximate the integrand over the domain $\mathcal{A}$. However, ADMB (using quadrature for integration) appeared to outperform early implementations of model-fitting functions that solely made use of R functions for optimisation (either `nlm()` or `optim()`) in terms of both speed and convergence. It was therefore selected as the optimiser for the developed R package.

Further effort could nevertheless be expended improving the approximation of the marginalisation over animals' locations. Any improvements would either improve accuracy of the approximation for a set computational budget, or decrease computation time for a set accuracy level. In particular, the current approach involves the use of evenly spaced mask points over the domain $\mathcal{A}$. While local integration can be used to reduce the number of mask points required for each individual, avenues that remain unexplored involve better selection of mask points, potentially using schemes that do not generate a uniform grid. For example, Gaussian quadrature rules evaluate the integrand at points that are instead roots of various types of polynomials (Legendre polynomials for Gauss-Legendre quadrature, Hermite polynomials for Gauss-Hermite quadrature, and so on), with each point given a set weighting. Such quadrature rules are typically more accurate than the simplistic rectangle rule.

An alternative possibility is to use an adaptive scheme. Evaluation of the integrand across points on a coarse grid spanning $\mathcal{A}$ could provide a low-resolution approximation of the function. This would allow for the identification of subsets of the domain within which the function should be further evaluated to increase its resolution; subsets of $\mathcal{A}$ in which the integrand is negligible would require no further function evaluations. This would

avoid the continued evaluation of the integrand at mask points that make inconsequential contributions to the approximation of the integral.

2.5.3 Concluding remarks

The unification of various established statistical methodologies (DS, MRDS, SECR, and CR) is satisfying from both academic and practical points of view. An umbrella model class that encompasses these is both theoretically elegant and suggests alternative ways of modelling previously collected data (e.g., Section 2.4.1). The development of the `admbsecr` package has great potential to accelerate the uptake of these methods, thus enhancing their impact upon—and visibility to—the wider ecological community.

The following three chapters provide further methodological development to this model class for various special-case scenarios.

Chapter 3

Cue-based SECR methods

3.1 Introduction

Acoustic SECR surveys typically collect cue-based detection data. That is, detectors detect individual cues, rather than animals, and it is not necessarily known which cues were emitted by which individuals; see Section 1.2.3 and Table 1.1. This makes estimation of animal density troublesome. It is not even known how many unique individuals were detected, and so there is no existing methodology that has been shown to appropriately estimate animal density from cue-based capture histories.

To see why, consider the basic likelihood presented in Chapter 1. Equations (1.10) and (1.11) show that the basic SECR likelihood can be derived as

$$L(\boldsymbol{\theta}; n, \boldsymbol{\Omega}) = f(n; \boldsymbol{\theta}) \int_{\mathcal{A}^n} f(\boldsymbol{\Omega}|n, \boldsymbol{X}; \boldsymbol{\gamma})\, f(\boldsymbol{X}|n; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{X}.$$

Two assumptions can be made to simplify the integral above: First, independence can be assumed between capture histories, conditional upon the locations they are emitted, giving

$$f(\boldsymbol{\Omega}|n, \boldsymbol{X}; \boldsymbol{\gamma}) = \prod_{i=1}^{n} f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma}).$$

Second, independence can be assumed between the locations of cues, giving

$$f(\boldsymbol{X}|n; \boldsymbol{\gamma}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i; \boldsymbol{\gamma}).$$

The combination of these two assumptions allows separability of the integral, giving rise to

the likelihood

$$L(\boldsymbol{\theta}; n, \boldsymbol{\Omega}) = f(n; \boldsymbol{\theta}) \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{\omega}_i | \boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}_i;$$

see Equation (1.12).

In many cases the first appears to be a reasonable assumption—given cues' locations, there is no reason to expect the capture history of one cue to provide information about that of another. For individual-based capture histories the second may be reasonable too, as the location of one individual may not provide any information about the locations of others. If, however, the capture histories are cue based, then this is unlikely to hold: the location of two cues emitted by the same animal are the same if animals are stationary, and are likely to be similar if they are not. Locations associated with cues from the same individual therefore cannot be considered independent.

Thus, strictly, the analysis of *B. acutorostrata* survey data in Chapter 2 is inappropriate: each capture history corresponds to a single surfacing of one whale, and it is possible that multiple surfacings of the same whale are detected, each occurring at similar locations. An analysis of *A. lightfooti* survey data using the methods proposed in Chapter 2 would suffer the same problem; individuals are stationary (or move a negligible distance) over the course of the survey, and so cues emitted by the same individual are associated with the same (or very similar) locations.

One possibility allowing analysis of such data is to derive the joint density $f(\boldsymbol{X}|n; \boldsymbol{\gamma})$. In order to achieve this it is necessary to model the dependence between cues. Assuming individuals can be identified from their cues, this is straightforward if individuals are stationary (as $\boldsymbol{x}_i = \boldsymbol{x}_{i'}$ if the ith and i'th cues were emitted by the same individual), but if they are not then the dependence must be accounted for by modelling individuals' intercue movements (see Section 7.1 for discussion on how this might be achieved). If individuals cannot be identified from their cues then modelling this dependence is more complicated still, as it is not clear whether or not any two cues were emitted from the same individual, thus the strength of the dependence is not clear.

Even in hypothetical cases in which the dependence *can* be specified (and so the joint density $f(\boldsymbol{X}|n; \boldsymbol{\gamma})$ is readily available) the integral is no longer separable. Rather than the evaluation of n two-dimensional integrals, likelihood computation involves the evaluation of a number of higher-dimensional integrals (if groups of capture histories can be considered independent of one another), or potentially a single $2n$-dimensional integral (if they cannot); both provide substantial computational obstacles. Although successfully implementing such an approach would hold great value, a more computationally efficient method

would nevertheless remain desirable.

While Efford, Dawson, and Borchers (2009) claimed to present a method of estimating density from acoustic signals, their approach assumed that each individual emitted exactly one cue over the course of the survey (in which case it is appropriate to assume independence between capture histories); naturally, this is incredibly unlikely to hold. Marques et al. (2012) and Martin et al. (2013) applied this method to data collected by an array of underwater hydrophones that detected calls of *B. acutorostrata*. As cue-based capture histories were analysed, they declared that their estimated parameter $\widehat{D}$ was *call* density rather than *individual* density (in their case, measured in calls per hour per 10 000 km^2). This is a sensible suggestion, given that DS methods involving the detection of cues (rather than individuals) have been shown to give rise to appropriate call density estimates (Buckland, 2006; Buckland et al., 2001; Hiby, 1985). However, these SECR models violate the independence assumption discussed above, and there has not yet been a formal investigation into the appropriateness of both point and variance cue-density estimators from SECR methods in this setting.

Moreover, it is typically *animal* density that is of interest. Existing DS methods convert from cue density to animal density by virtue of collecting independent cue-rate data (e.g., by using a separate survey to track down a number of individuals, and recording how often they produce cues; Buckland, 2006; Buckland et al., 2001; Hiby, 1985). The quotient of the cue-density and the cue-rate estimates (i.e., the average cue rate across surveyed individuals on the independent survey) provides an estimate of animal density. Again, this approach has not yet been tested for SECR methods.

This chapter presents a method of estimating animal abundance from cue-based capture histories (Section 3.2). It is computationally tractable (in that it does not rely on the evaluation of high-dimensional integrals) and does not rely on individuals being identifiable from their cues. Standard SECR models (that assume independence between locations) applied to data with cues as the unit of detection are found to estimate cue density with negligible bias despite the violated assumption. Likelihood-based variance estimates are negatively biased, however, resulting in CIs that do not reach their nominal coverage levels. Collection of independent cue-rate data allows for estimation of animal density, as well as the application of a novel variance estimation method. A simulation study is used to show that the proposed point and variance estimators are negligibly biased, and result in interval estimators that reach their nominal coverage level (Section 3.4). This approach is used to estimate density of *A. lightfooti*, an anuran endemic to the Western Cape province of South Africa at a particular site at a particular time (Section 3.5.2). Additionally, it is used to

estimate variation of *A. lightfooti* calling density across space, time, and over environmental covariates such as rainfall and temperature (Section 3.5.3).

The methodology developed in this chapter formed the basis of a publication (Stevenson et al., 2015), which appeared in *Methods in Ecology and Evolution*, November 2014. The application of these methods to *A. lightfooti* survey data found in this chapter also appears in a further manuscript (Measey, Stevenson, Scott, Altwegg, & Borchers, in submission), which has recently been submitted to the *Journal of Applied Ecology*.

3.2 Methodology

Here, Equation (2.1) is used as a density estimator. However, as i now indexes cues (rather than individuals), D is assumed to refer to cue (rather than animal) density. Common forms of auxiliary data that can be employed in this scenario include signal strengths and TOAs.

A simple random sample of n_r individuals from the population are monitored independent of the main survey, but at the same time and location. A cue rate (in cues per unit time) is observed from each. Let r_i be the cue rate observed from the ith monitored individual, $\boldsymbol{r}$ hold all observed cue rates, and $\widehat{\mu}_r = \sum_{i=1}^{n_r} r_i/n_r$ provide an estimator for the mean population cue rate, μ_r.

An estimator of animal density is therefore

$$\widehat{D}_a = \frac{\widehat{D}}{\widehat{\mu}_r}. \tag{3.1}$$

The unit of the numerator is cues per unit time per unit area, while that of the denominator is cues per individual per unit time. The quotient of the two thus has the unit individuals per unit area.

While this estimator may be intuitive, its properties are not immediately clear. For example, if $\widehat{D}$ and $\widehat{\mu}_r$ are unbiased estimators of D and μ_r, respectively, then it does *not* follow that $\widehat{D}_a$ is also unbiased; the expectation of a quotient of random variables is not equivalent to the quotient of the respective expectations, and so, in general, $\mathrm{E}(\widehat{D}_a) = \mathrm{E}(\widehat{D}/\widehat{\mu}_r) \neq \mathrm{E}(\widehat{D})/\mathrm{E}(\widehat{\mu}_r)$.

Here, the estimator shown in Equation (3.1) is given further theoretical justification. It is shown that, if the estimator $\widehat{D}$ has asymptotic unbiasedness and normality as $t \to \infty$ (thus as $n \to \infty$), then the properties of asymptotic unbiasedness and normality also hold for $\widehat{D}_a$; that is, if the estimator $\widehat{D} \xrightarrow{d} \widehat{D}_* \sim \mathrm{N}(D, \mathrm{Var}(\widehat{D}))$ as $t \to \infty$, where $\widehat{D}_*$ is a

random variable with the asymptotic distribution of $\widehat{D}$, then the estimator $\widehat{D}_a = \widehat{D}/\widehat{\mu}_r$ has asymptotic unbiasedness and normality.

The estimator $\widehat{\mu}_r$ is the mean of a simple random sample from the population of animals; thus, by the weak law of large numbers, $\widehat{\mu}_r \xrightarrow{p} \mu_r$ as $n_r \to \infty$.

Therefore, as $t \to \infty$ and $n_r \to \infty$,

$$\widehat{D}_a = \frac{\widehat{D}}{\widehat{\mu}_r} \xrightarrow{d} \frac{\widehat{D}_*}{\mu_r},$$

due to Slutsky's theorem. From the condition of asymptotic unbiasedness and normality of $\widehat{D}$ above,

$$\widehat{D}_a \xrightarrow{d} \frac{\widehat{D}_*}{\mu_r} \sim \mathrm{N}\left(\frac{D}{\mu_r}, \frac{\mathrm{Var}(\widehat{D})}{\mu_r^2}\right). \tag{3.2}$$

Thus, asymptotically (as both $t \to \infty$ and $n_r \to \infty$),

$$\mathrm{E}(\widehat{D}_a) = \frac{D}{\mu_r} = D_a.$$

If $\widehat{D}$ were an ML estimator, then the condition above (i.e., its asymptotic unbiasedness and normality) would be met directly from standard ML theory. However, the objective function being maximised (Equation (2.3)) for its estimation (Equation (2.1)) is the likelihood of a model that does not befit the data-generating process as it assumes independence between cue locations. Therefore, one cannot directly project ML estimator properties onto $\widehat{D}$. The appropriateness of the above condition is instead assessed via simulation in Section 3.4.2.

For the same reason, ML variance estimators should not be used for model parameters in this context. Typically, in cases where data violate a model's independence assumption due to positive correlation across sampling units, variance estimates are underestimated. This causes CIs to have a true coverage lower than their nominal levels.

Here an alternative is proposed. It is a simulation method that uses parameter estimates ($\widehat{D}$, $\widehat{D}_a$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{\psi}}$, and $\widehat{\mu}_r$) to generate numerous data sets with the appropriate cue-location dependence. The parameter estimates obtained from each allow inference of the estimators' properties under cue-location dependence, thus giving rise to appropriate variance estimates. This approach is similar to a parametric bootstrap procedure, only the statistical model used to derive estimators is different to that used for data simulation.

For this approach it is also necessary to estimate the distribution of population cue rates. This can be done either parametrically—for example, by assuming $\boldsymbol{r}$ has CDF $F(r; \boldsymbol{\zeta})$

and estimating $\boldsymbol{\zeta}$ in some way—or nonparametrically—for example, by using the empirical distribution function (EDF) to estimate $F(r; \boldsymbol{\zeta})$.

Here the superscript * is used to denote either simulated data, or estimates obtained from simulated data. The following procedure is proposed:

1. Simulate animal locations within $\mathcal{A}$ as a realisation of a Poisson point process with homogeneous intensity $\widehat{D}_a$.
2. Determine the number of cues made by each individual by simulating from the estimated distribution of $\boldsymbol{r}$ (e.g., either some parametrically estimated distribution or the EDF).
3. Construct $\boldsymbol{X}^*$. The location of each cue is given by the location of the individual that produced it.
4. Generate $\boldsymbol{\Omega}^*$ via simulation of detections, with detection probabilities given by $g(d_j(\boldsymbol{x}); \widehat{\boldsymbol{\gamma}})$. Remove entries in $\boldsymbol{X}^*$ and $\boldsymbol{\Omega}^*$ that are associated with cues that were not detected by any detector.
5. Generate $\boldsymbol{Y}^*$ (for any auxiliary data collected during the survey) via simulation from the PDF $f(\boldsymbol{Y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})$ (Equation (2.4)).
6. Calculate $\widehat{D}^*$, $\widehat{\boldsymbol{\gamma}}^*$, and $\widehat{\boldsymbol{\psi}}^*$ from $\boldsymbol{\Omega}^*$ and $\boldsymbol{Y}^*$ (Equation (2.1)).
7. Generate $\boldsymbol{r}^*$ via simulation of n_r independent and identically distributed (IID) random variables from the estimated distribution of $\boldsymbol{r}$ (e.g., either from $F(r; \widehat{\boldsymbol{\zeta}})$ or its EDF).
8. Calculate $\widehat{\mu}_r^* = \sum_{i=1}^{n_r^*} r_i^* / n_r^*$
9. Calculate $\widehat{D}_a^* = \widehat{D}^* / \widehat{\mu}_r^*$.
10. Repeat the above steps n_b times, saving $\widehat{\boldsymbol{\theta}}^*$ on each occasion.

One particular point to note is that, in Step 2, the simulated cue-rate data may provide non-integer numbers of cues emitted by individuals. In this case, let $r^{*\prime}$ be the non-integer value generated. The resultant number of cues simulated is then given by

$$r^* = \begin{cases} \lceil r^{*\prime} \rceil & \text{with probability } r^{*\prime} - \lfloor r^{*\prime} \rfloor, \\ \lfloor r^{*\prime} \rfloor & \text{with probability } 1 - (r^{*\prime} - \lfloor r^{*\prime} \rfloor), \end{cases}$$

and so $r^{*\prime}$ is rounded up or down at random to give r^*, such that $\mathrm{E}(r^*) = r^{*\prime}$.

The n_b sets of saved parameter estimates can then be treated as though they were generated from a bootstrap procedure; see Davison and Hinkley (1997) for a detailed account of what can be achieved. In particular, standard deviations of estimates obtained from simulated data provide standard errors, and biases can be estimated by subtracting the original parameter estimate from the mean of the estimates from simulated data; a bias adjustment can then be carried out by subtracting this bias from the original estimate. A variety of bootstrap CI method exist, some of which are summarised in Section 2.3.3.

One final point to note is that the objective function maximised for point estimation is not the 'true' likelihood for the model being fitted; thus likelihood-based information criteria (such as AIC) should no longer be used for model selection. See Section 3.6.2 for discussion on this point, and a potential alternative.

3.3 Implementation in `admbsecr`

The above methods can be carried out using the `admbsecr` package. The `call.freqs` argument of `admbsecr()` accepts a vector of cue-rate data that have been collected independent of the main survey.

If this argument has been provided, then (i) the animal density estimate, D_a, is saved and returned along with all other parameters upon calling of functions such as `coef()` and `summary()`; and (ii) the `admbsecr()` function does not estimate the Hessian—measures of parameter uncertainty (i.e., standard errors and CIs) cannot be calculated and are not returned[1]. Instead, the `boot.admbsecr()` function (Section 2.3.3) can be used to carry out the bootstrap-like procedure detailed above. This is automatically invoked (rather than a standard parametric bootstrap procedure) if the fitted model object was created using a call to `admbsecr()` that provided cue-rate data via the `call.freqs` argument. The standard utility functions (e.g., `stdEr()`, `confint()`, `summary()`, and `vcov()`; Section 2.3.2) then allow inference based on the results of this procedure.

3.4 Simulation studies

3.4.1 Estimator properties

A simulation study was carried out in order to evaluate the performance of both the point and variance estimators described above. In total, 1 000 data sets were simulated following

[1]This default behaviour can be overridden by setting the optional argument `hess` to `TRUE`, but, in this case, measures of parameter estimate uncertainty should probably not be trusted.

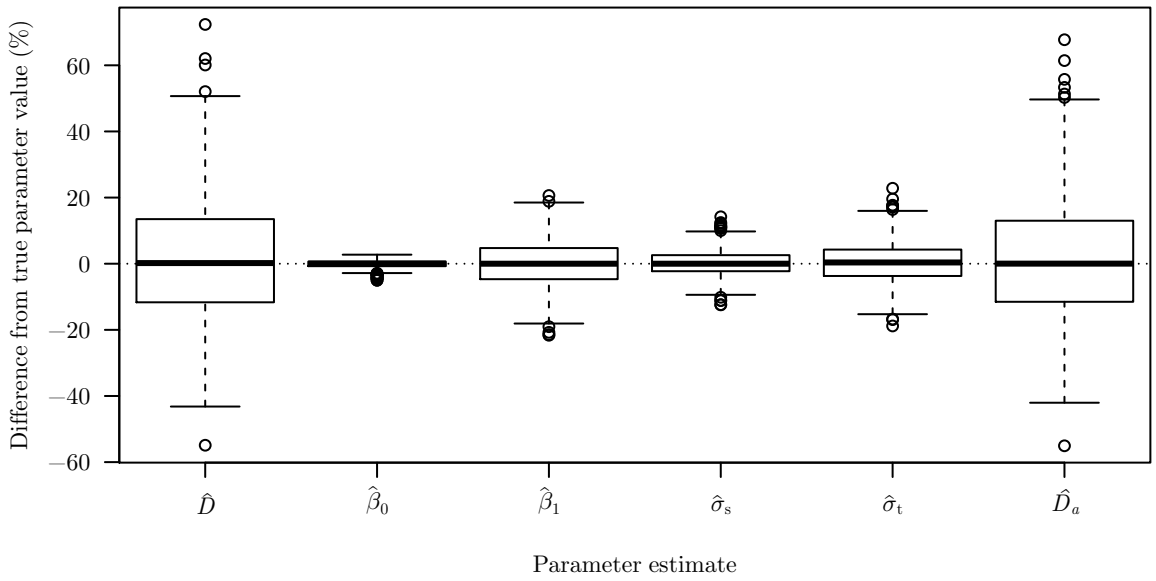


Figure 3.1 *Boxplot showing point estimates for parameters of interest from each of the simulated data sets. These are given as percentage differences from the true parameter value from which the data were generated.*

a similar scheme to that outlined for the bootstrap-like procedure: animal locations were generated across a survey region with homogeneous intensity 366.08 individuals per hectare, each of which producing cues at an average rate of 6.77 cues per minute (with a standard deviation of 0.37 cues per minute). Cues emitted by the same individual were simulated as originating from the same location. Cue detections were simulated based on a signal strength detection function with an identity link, with source strength $\beta_0 = 156.57$ units, signal loss $\beta_1 = 2.67$ units m^{-1}, and measurement error $\sigma_s = 11.50$ units. The error in measured TOAs was set at $\sigma_t = 0.002$ s. Surveys were simulated at a duration of $t = 25$ s. These parameter values match the estimates obtained from the *A. lightfooti* survey data, and the layout of the detector array used for data simulation also mirrored this application (see Section 3.5.3 and Figure 3.4).

All point estimators of interest were shown to be (at most) negligibly biased (Figure 3.1, Table 3.1); thus, it is apparent that (i) the cue density estimator, $\widehat{D}$, performs well despite being obtained from a likelihood that corresponds to a model that has a key assumption violated by the data; and (ii) it is appropriate to use an animal density estimate, $\widehat{D}_a$, that is calculated from the quotient of cue density and cue rate estimates (Equation (3.1)).

As expected, CIs based on standard ML theory were generally too narrow, and did not reach their nominal coverage levels. The simulation procedure described in Section 3.2 performed well. In particular, the normal and percentile bootstrap CI methods showed

Table 3.1 *Estimated biases, variances, and root mean square errors from the simulation study for estimators of interest. All are given as percentages of the true underlying parameter values used to generate the data.*

Estimator	Bias (%)	SD (%)	RMSE (%)
$\widehat{D}$	1.14	18.22	18.26
$\widehat{\beta}_0$	−0.11	1.12	1.13
$\widehat{\beta}_1$	−0.01	6.71	6.71
$\widehat{\sigma}_s$	0.20	3.87	3.88
$\widehat{\sigma}_t$	0.35	5.97	5.98
$\widehat{D}_a$	1.13	18.28	18.32

Table 3.2 *Estimated coverage for a variety of CI methods. The 'Naïve' method is based on standard errors calculated from the inverse of the observed Fisher information matrix, calculated from the likelihood that assumes independence between cue locations. This method cannot compute a CI for the parameter D_a as it does not appear in this likelihood. Nominal CI coverage was set at 95% in all cases.*

	Parameter					
CI method	D	β_0	β_1	σ_s	σ_t	D_a
Normal	0.941	0.957	0.959	0.949	0.953	0.942
Basic	0.927	0.950	0.959	0.949	0.935	0.924
Percentile	0.946	0.920	0.955	0.947	0.972	0.942
Naïve	0.729	0.863	0.919	0.946	0.941	—

appropriate coverage for cue and animal density parameters (D and D_a), with the basic method reporting slight undercoverage. Interestingly, CIs for auxiliary information measurement error parameters (σ_s and σ_t) appeared appropriate for all methods, including the likelihood-based approach (Table 3.2). The average standard error for each parameter (calculated via the simulation method) matched closely with the standard deviations of the respective 1 000 parameter estimates obtained from the simulated data sets, further indicating that the methodology developed here gives rise to unbiased variance estimators.

3.4.2 Asymptotic normality

A simulation study was carried out in order to determine the asymptotic distribution of the estimator $\widehat{D}$. Data were generated as above (Section 3.4.1), but for surveys of varying lengths (1.56, 3.13, 6.25, and 12.5 s). The results from the simulation study in Section 3.4.1 were based on surveys of 25 s in length, and were reused here. Cue density estimators converged to normality as survey length increased (Figure 3.2), with estimator variance a

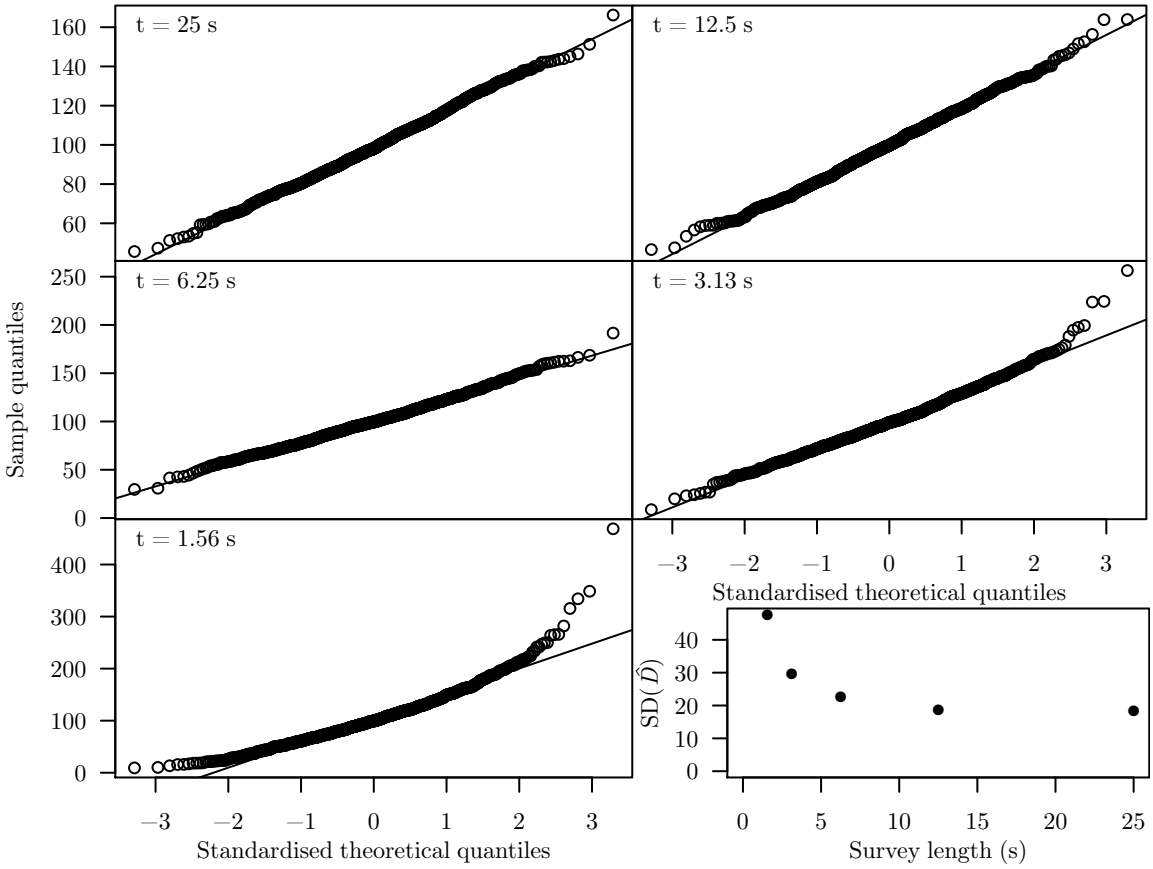


Figure 3.2 *QQ plots of estimated cue densities from simulated data for various survey lengths. The estimates converge on normality as survey length increases. Inset: survey length plotted against standard deviations of these estimates, showing decreasing estimator variance as t increases.*

decreasing function of t; in turn, this implies asymptotic normality of the estimator $\widehat{D}_a$ (Section 3.2, Equation (3.2)).

3.5 Applications to *A. lightfooti* survey data

The genus *Arthroleptella* (family Pyxicephalidae) comprises seven species of frogs endemic to the Western Cape province of South Africa. They inhabit mossy seepages in montane fynbos areas (Channing, 2004) and many species are confined to a single mountaintop—they are highly range-restricted. Moreover, this habitat has a natural fire cycle; however, invasive woody species (such as pine and Australian acacia) increase both fire-event frequency and fire temperatures (Kraaij & van Wilgen, 2014). As a result, most species of the genus appear on the IUCN red list (Measey, 2011).

In particular, *A. lightfooti* is listed as 'Near Threatened'. Estimation of their abundance and density is necessary to monitor their populations and better understand their ecology. Individuals are tiny (females can reach a snout–vent length of 22 mm; males are smaller still) and are cryptically coloured. They are thus prohibitively time-consuming to find—this takes approximately 3–4 person-hours per individual—and so traditional monitoring surveys involving the physical capture and handling of individuals are not feasible.

Aestivation takes place over austral summer, while an active, breeding period occurs over the rainy season (April–December), during which time choruses of male advertisement calls develop (Channing, 2004). These calls can be distinguished from other species of *Arthroleptella* (e.g., Turner & Channing, 2008; Turner, De Villiers, Dawood, & Channing, 2004), although they are not sympatric with any other species of the genus on the Cape peninsula (Channing, 2004). Calling *A. lightfooti* are therefore readily detectable via auditory monitoring, and can be reliably identified from other anuran species. Individuals, however, are not easily acoustically identifiable—at present it cannot be determined with certainty whether or not any two detected calls were emitted by the same animal.

Manual calling surveys (MCSs) and automated recording system (ARS) are two existing anuran auditory monitoring methodologies. The former involves surveyors listening to a chorus, and then subjectively estimating the number of calling males. The latter deploys an automated recorder to collect a permanent recording of the acoustic data, and is typically used to determine site occupancy (although the amplitude of a chorus may be used as a call-abundance index). These approaches, however, are not without their problems: (i) MCSs are inherently subjective—while this is not ideal in itself, it also means that there are no corresponding measures of estimate uncertainty; and (ii) neither method formally accounts for the ESA—thus density (rather than abundance) estimates are not available, and any change in abundance could either be due to a change in acoustic detectability (and so a change in ESA) or a genuine change in call or animal density. Indeed, Dorcas, Price, Walls, and Barichivich (2009) conclude that existing auditory monitoring methods have a limited ability to estimate anuran abundance and density.

The methodology described earlier in this chapter overcomes these issues. It makes objective estimates of both abundance and density available from acoustic detection data collected by a fixed array of detectors, and recall that SECR methods explicitly estimate and account for a survey's ESA (Section 1.2.1, Equation (1.6)). Importantly, the estimators proposed here should give asymptotically unbiased estimates of animal density, even under changes in detectability. This allows for the investigation of changes in animal or call density across time and space, greatly enhancing the potential of auditory monitoring surveys of

A. lightfooti and other species of the genus.

3.5.1 Acoustic surveys of *A. lightfooti*

Throughout May–September 2012, acoustic surveys of *A. lightfooti* were carried out across three sites on the Steenberg Plateau in Silvermine Nature Reserve, Table Mountain National Park, on South Africa's Cape peninsula. These have GPS coordinates of S 34°06′03.5″ E 18°26′55.2″, S 34°05′51.0″ E 18°26′56.8″, and S 34°05′57.7″ E 18°27′03.8″.

In total, there were seventeen visits to the study region. On each visit two sites were usually surveyed, as time constraints did not permit the surveying of all three. Site 1 was always surveyed, along with either Site 2 or 3 on all but three occasions. For each survey, six Audio-Technica AT8004 handheld, omnidirectional, dynamic microphones were connected to a DR-680 8-Track portable field audio recorder with Hosa Technology STX-350F Professional 1/4 inch TRS male to XLR female cables. Each microphone was placed in a microphone holder, fastened atop a wooden dowel that stood at a height of 1 m. To ensure that microphones were located in the same positions at each return visit, the dowels were placed in plastic tubes that remained in the ground throughout the entire study period. The microphone layouts across the three sites can be seen in Figure 3.6. Each survey recorded vocalising *A. lightfooti* for forty minutes. The first ten minutes were discarded in case calling behaviour was affected by disturbance during the installment of the microphone array, and the area around the site was vacated while recording took place. This survey period was sufficiently short for it to be appropriate to assume that calling *A. lightfooti* remained stationary throughout—males typically call at (or near) the site that a female has laid its eggs (Channing, 2004).

Advertisement calls of *A. lightfooti* are chirps of three short pulses, with an emphasised frequency of 3 754 Hz (Turner & Channing, 2008). These recordings were processed with PAMGuard (version 1.11.00 BETA; Gillespie et al., 2008)—open-source software designed for the processing of cetacean acoustic detection data. PAMGuard scanned the recordings from each microphone to identify received acoustic signals that matched properties of a characteristic *A. lightfooti* call. The data returned by PAMGuard are in the form of a data frame, where each column corresponds to a detection of a call at a particular microphone. Various variables are held within this; only (i) the precise time (in seconds) from the start of the survey that the beginning of a detected call was recorded and (ii) the amplitude (i.e., strength) of the received signal were used here. These provide the TOA and signal strength for each detection, respectively.

In order to generate the required capture histories, $\boldsymbol{\Omega}$, it is necessary to determine which

identified calls across different microphones are due to recordings of the very same call. This was achieved purely from the precise TOAs recorded: if two detections were at microphones separated by distance d, then they were assumed to have been due to recordings of the same source if the difference in their respective TOAs was less than d/s s. That is, if the TOAs were similar enough for it to be possible that the detections were of the same call, then this was assumed to be the case. This approach should never misallocate two detections of the same call to two separate calls; however, there is potential for detections of calls emitted at similar times to be misallocated as having the same source. While this approach to constructing $\boldsymbol{\Omega}$ is therefore not suitable for all species, male *A. lightfooti* tend to call in turn with sufficient waiting times between individuals (G. J. Measey, personal observation), and so it appears appropriate here.

The `admbsecr` package exports the function `convert.pamguard()`; this takes this data frame as its main argument, and returns a capture history object in the appropriate format for the `capt` argument of the `admbsecr()` function (see Section 2.3.1). In order to do this it allocates detections to calls using the above procedure. Other arguments of this function include (i) `mics`, providing the microphone locations; (ii) `time.range`, to allow truncation of the data to a particular time range (provided in seconds from the beginning of the survey); and (iii) `sound.speed`, the speed of sound, s.

3.5.2 Estimation of *A. lightfooti* density

The first surveys were conducted on 16 May 2012 at Sites 1 and 3. Density of calling *A. lightfooti* males at Site 1 is estimated here in order to demonstrate the approach described in Section 3.2. Although no cue-rate data were collected simultaneously to the acoustic survey, they were collected on a different day, but at a similar time of the year. These are used in this section in order to illustrate the use of the estimation method.

This forty-minute survey detected a total of 12 064 individual calls. Unfortunately, the sheer abundance of data collected is such that not all can be used for analysis due to considerations of computing time—particularly when one considers that the simulation method (Section 3.2) requires the repeated refitting of the same model to newly generated data. It was therefore necessary to truncate these data, and, to this end, only the first 25 s were used for analysis. This corresponded to a total of 225 detected calls, a far more manageable number.

These data are exported by the `admbsecr` package, and are found in the `lightfooti` object; this is available upon loading of the package. It is a list containing six named components:

Table 3.3 *Point estimates, standard errors, 95% CIs, and estimated biases for parameters of interest from the* A. lightfooti *data. The parameter D is measured in calls per hectare per second, σ_t in milliseconds, and D_a in calling males per hectare. Signal strengths were measured on an arbitrary scale. The CIs were calculated using the percentile method.*

			95% CI		
Parameter	Estimate	SE (% MCE)	Lower	Upper	Bias (%)
D	99.15	17.39 (0.70)	67.09	135.06	0.59
β_0	156.57	1.81 (1.00)	152.12	159.32	−0.14
β_1	2.67	0.18 (0.77)	2.31	3.02	−0.22
σ_s	11.50	0.44 (0.78)	10.66	12.40	−0.07
σ_t	1.96	0.12 (0.76)	1.73	2.20	0.60
D_a	366.08	64.63 (0.70)	247.27	498.76	0.62

- `capt`: an object containing capture history, signal strength, and TOA information for each detected call, $\boldsymbol{\Omega}$ and $\boldsymbol{Y}$; this is suitable for use as the `capt` argument of `admbsecr()`.
- `traps`: an object containing the locations of the detectors; this is suitable for use as the `traps` argument of `admbsecr()`
- `mask`: an object providing a suitable habitat mask, for use as the `mask` argument of `admbsecr()`.
- `cutoff`: the threshold signal strength, c.
- `freqs`: the independently observed call rate data, $\boldsymbol{r}$; this is suitable for use as the `call.freqs` argument of `admbsecr()`.
- `dets`: the data frame returned by PAMGuard.

The object component `capt` can therefore be created by calling

```
capt <- convert.pamguard(lightfooti$dets, mics = lightfooti$traps)
```

For all code displayed throughout the remainder of this section, assume that components of the `lightfooti` object can be accessed by specifying only their names. This, therefore, allows the replacement of the unwieldy `lightfooti$capt` (above) with `capt`. This is equivalent to running `attach(lightfooti)` prior to execution of any following code chunks.

Carrying out the estimation procedure can be achieved using

```
lightfooti.fit <- admbsecr(capt = capt, traps = traps, mask = mask,
                           ss.opts = list(cutoff = cutoff),
                           call.freqs = freqs)
```

Fitting either the log-link or spherical-spreading signal strength detection functions simply requires inserting the component `ss.link = "log"` or `ss.link = "spherical"` into the `ss.opts` list, respectively. With these data, the objective function being maximised (Equation (2.2)) was largest with the identity link, and so this was selected as the best-fitting detection function[2]. The simulation procedure (Section 3.2) was then implemented[3] using

```
lightfooti.boot.fit <- boot.admbsecr(lightfooti.fit, N = 10000)
```

Parameter estimates, CIs, standard errors, and estimated biases are shown in Table 3.3. The estimated detection function is shown in Figure 3.3, while the effects of incorporating signal strength and TOA data into location estimates are shown in Figure 3.4.

Estimates were relatively precise—both animal and call density parameters were associated with coefficients of variation (CVs) under 20%—and biases were estimated to be negligible. Additionally, note that the number of simulated data sets, n_b, was set at 10 000 as this was sufficiently large to reduce the relative MCE (Koehler et al., 2009) in the standard errors of all parameters to below 1%. Once more, incorporation of auxiliary information substantially improved location estimates. The estimated detection function matched expectations—it was hypothesised that a call emitted directly below a microphone would almost certainly be detected, and its range (with very few detections beyond 20 m) was also consistent with expectations.

3.5.3 Investigation of *A. lightfooti* call density variation

It was of interest to investigate changes in *A. lightfooti* call density over an entire calling season. Acoustic monitoring is best carried out when the largest proportion of males are actively calling, and in order to identify this time it is necessary to ascertain when call density is at its highest. Collection of data at three separate sites also allowed exploration into the degree of spatial variation in call density.

[2]Recall that likelihood-based information criteria should not be used for model selection (Section 3.2). In this case, however, all models have the same number of parameters; there is no compromise to make between model fit and complexity, and so that with the largest maximised objective function can be considered to be best supported by the data at hand.

[3]This is best carried out utilising the parallel processing capabilities of `boot.admbsecr()` and perhaps left to run on a server.

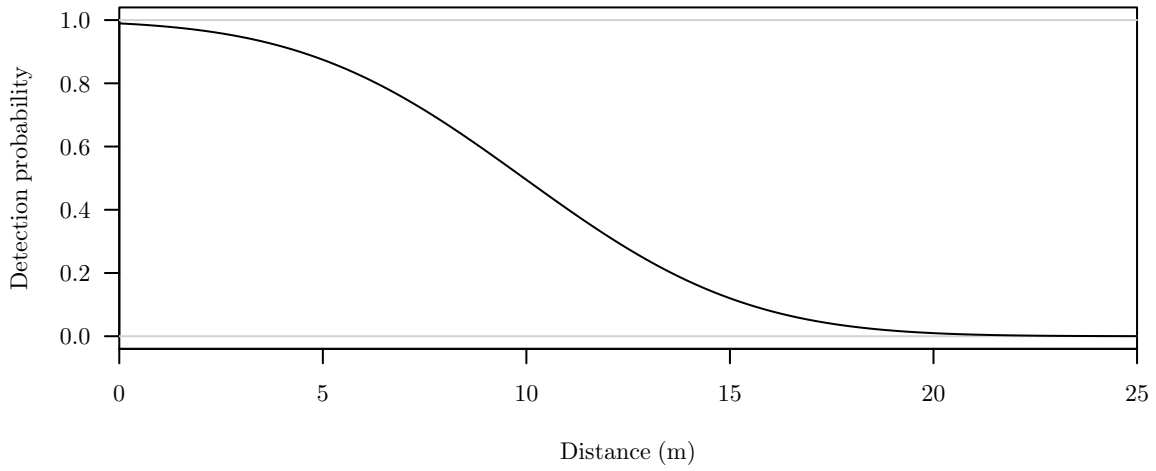


Figure 3.3 *The estimated detection function from the* A. lightfooti *data.*

Furthermore, calling behaviour of other anuran species has been found to vary with temperature and rainfall (Hauselberger & Alford, 2005; Kirlin, Gooch, Price, & Dorcas, 2006; Murphy, 2003; Navas, 1996; Oseen & Wassersug, 2002; Saenz, Fitzgerald, Baum, & Conner, 2006; Weir, Royle, & Jung, 2005). In order to explore relationships between *A. lightfooti* call density and both of these environmental variables, daily measurements were taken of both rainfall (by park rangers from a rain gauge located approximately three kilometres from the three sites) and temperature (by a temperature logger placed at Site 1; this collected both ground and air temperatures). It was hypothesised that *A. lightfooti* calling behaviour may be related to rainfall up to two days prior.

As per Section 3.5.2, each survey collected more data than it was possible to analyse with a single SECR model; however, discarding a vast majority of the data was not desirable. As a compromise, ten regular, one-minute subsets of the recordings were taken, and a call-density estimate was obtained from each; an overall call-density estimate for the survey was then obtained by averaging over these. Thus, ten minutes of data were analysed from each survey in a way that is far computationally cheaper than attempting to fit a single SECR model to ten minutes' worth of data.

Relationships between these call-density estimates (as the response variable) and time, site, rainfall, and temperature (as explanatory variables) were investigated with a second stage of modelling. Exploratory plots indicated that there were differences in call density between sites, and so these were modelled as fixed effects. The relationship between call density and time appeared to be nonmonotonic: call densities were lowest in May, increased towards a peak in late July, before declining again towards the end of the season (Figure 3.5); thus a quadratic temporal effect was incorporated. During an initial, exploratory

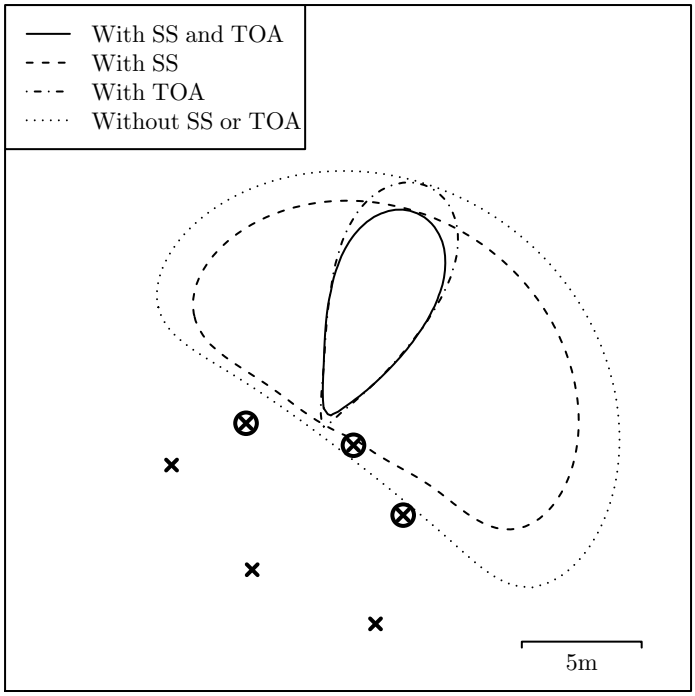


Figure 3.4 *Estimated conditional PDFs of a detected call's location, with various combinations of auxiliary information. Crosses represent microphone locations, while those that are circled detected this particular call. It is estimated that the call was located within each contour with a probability of 0.9.*

modelling process, there was strong support for separately estimated temporal effects (i.e., an interaction between time and site).

It was noted that the temporal effect appeared to be multiplicative in nature, whereby the higher a site's call density, the larger the absolute change in abundance across time. For example, across all sites, call densities in the middle of the season appeared to be roughly three times as large as those at the beginning of the season (see Figure 3.5). A log transformation was thus applied to the call densities to allow the fitting of a multiplicative temporal relationship, and following this there was no longer any suggestion that interaction effects were required. This allowed a far more parsimonious model, with only four parameters required to explain the spatial and temporal effects (two fixed site effects, and two for the quadratic temporal effect) rather than eight for the previous model (two fixed site effects, then two for a quadratic temporal effect at each site).

A side effect of the log transformation was that error variance appeared to decrease with increasing expected call density. To account for this, generalised least squares procedure (rather than ordinary least squares) was used to fit the linear model in such a way that

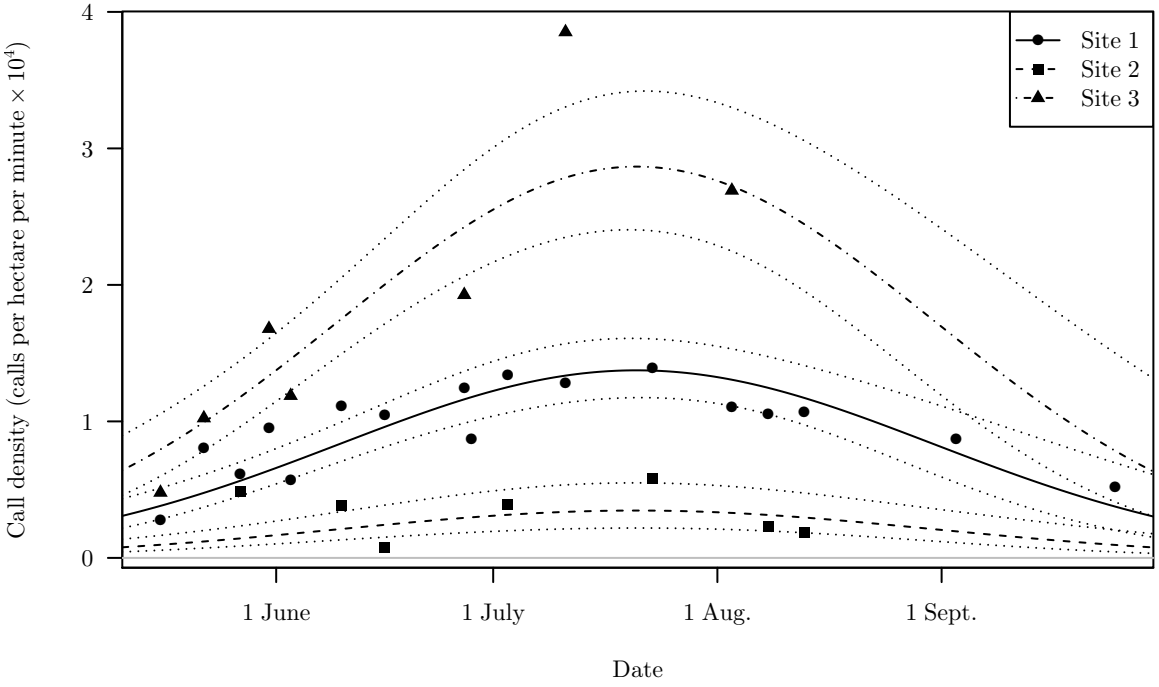


Figure 3.5 *Estimated* A. lightfooti *call densities across the three sites over the duration of the calling season. Solid, dashed, and dot-dashed lines provide the estimated expected call densities at Sites 1, 2, and 3, respectively; these are given by the linear model fitted by generalised least squares to the log-transformed call density estimates. Dotted lines provide lower and upper bounds of* 95% *CIs for these expectations.*

error variance depended on fitted value subject to a power-variance function structure. This modelling decision was strongly supported by AIC.

Additionally, it was hypothesised that surveys carried out on the same day may be similar in some way (e.g., due to comparable, unmodelled environmental conditions), and so random day effects were included in the model; however, a simulation-based, exact likelihood-ratio test (Crainiceanu & Ruppert, 2003) provided no evidence against the hypothesis that the associated variance component was zero, and so these were subsequently removed in the interests of parsimony.

All combinations of environmental covariates (chosen from site, time, rainfall up to two days prior, and temperature) were then fitted under this framework. Models were then compared using AICc (the 'corrected' AIC); see Table 3.4. The model best supported by the data included only site and temporal effects; thus there was weak support for models that relate call density to either rainfall or temperature. There was, however, substantial evidence to suggest that call densities varied across both time and space (Figure 3.5).

Interestingly, there was considerable variability in detection across the three sites (Fig-

Table 3.4 *The best ten linear models (based on AICc) fitted by generalised least squares. All included fixed site effects (two parameters), a quadratic temporal effect (two parameters), an intercept parameter, and two parameters to explain error variance. The best-fitting model did not include any other environmental covariates. The 'ΔAICc' column gives the difference in AICc from this model. The 'AICc weight' column provides relative weightings of importance across all models.*

Additional variables	No. parameters	ΔAICc	AICc weight
None	7	0.000	0.305
Air temperature	8	1.013	0.184
Ground temperature	8	3.087	0.065
Current rainfall	8	3.336	0.057
Rainfall one day prior	8	3.403	0.056
Rainfall two days prior	8	3.503	0.053
Air temperature Current rainfall	9	3.919	0.043
Air temperature Rainfall two days prior	9	4.748	0.028
Air temperature Ground temperature	9	5.012	0.025
Air temperature Rainfall one day prior	9	5.039	0.025

ure 3.6); this, in turn, resulted in substantial between-site variability in ESA. Microphones at Site 2 appeared capable of detecting calls from the greatest distances, and—from the two remaining locations—those at Site 3 appeared to better detect calls emitted at short distances. However, in general, detection functions estimated at Site 1 declined more gradually. It is plausible that this variation was related to the vegetation at each of the three sites: Site 3 had tall, dense vegetation that calls must penetrate in order be detectable at the microphones; Site 2 had the most open vegetation; and Site 1 was intermediate, with dense but low vegetation.

3.6 Discussion

The methods presented in this chapter are the first that allow for appropriate animal density estimation via acoustic SECR models in situations that do not allow identifiability of individuals; the framework described therefore provides a generalisation of sizeable importance,

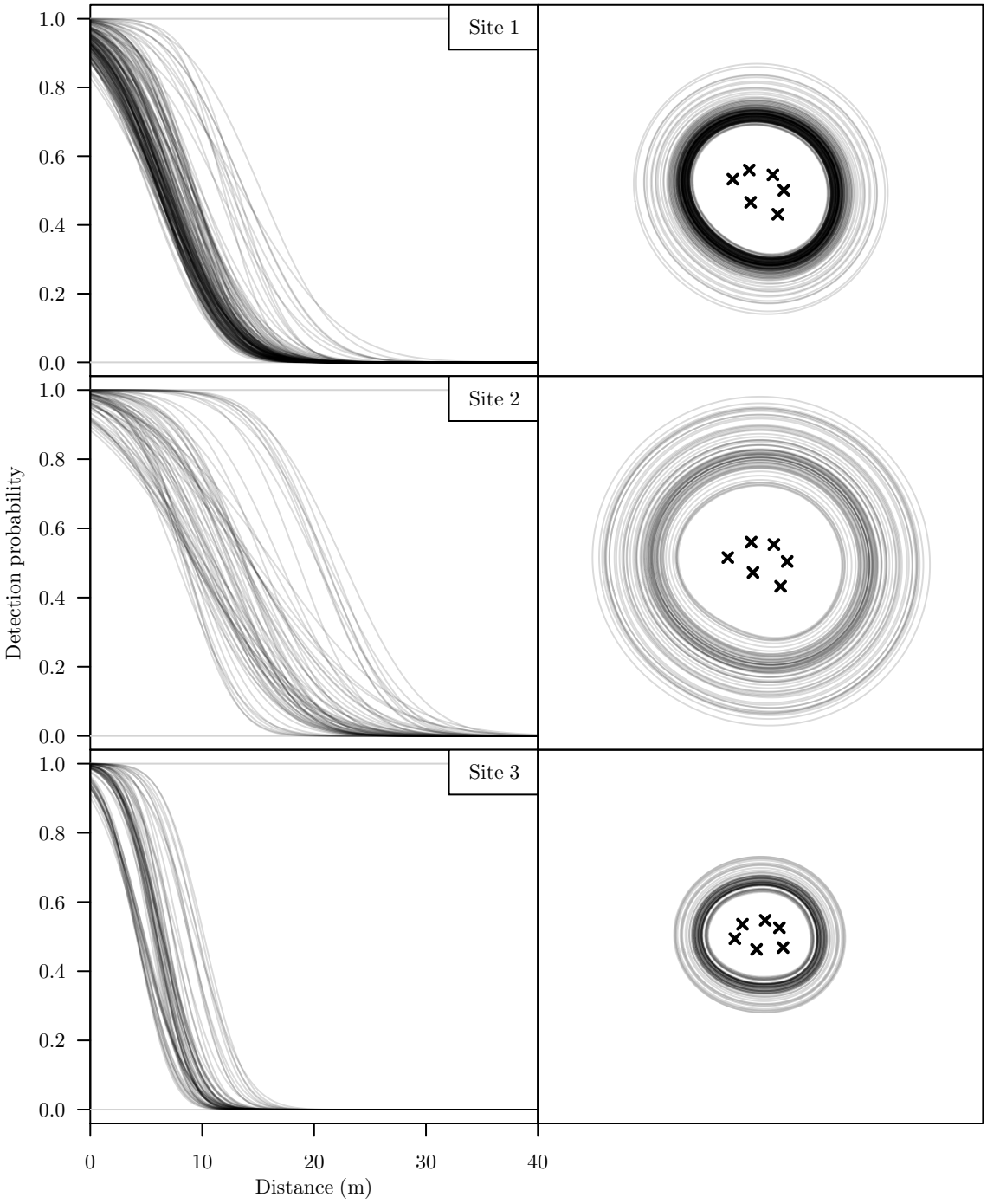


Figure 3.6 *Estimated detection functions from each one-minute sample (left column), along with depictions of the associated estimated ESAs (right column). The contours plotted in the right column are such that they encapsulate an area equal to the estimated ESA; thus it is estimated that, for a particular one-minute sample, the number of calls originating within the associated contour that were missed is equal to the number of calls emitted outwith the contour that were detected.*

allowing the use of SECR with data from a considerably wider variety of acoustic surveys. The simulation studies (Section 3.4) show that the methodology developed gives rise to appropriate estimators, allowing precise animal density estimation from short surveys that are associated with minimal fieldwork costs.

3.6.1 Collection of cue-rate data

The primary disadvantage lies in the requirement of collecting independent cue-rate data. In particular, cue rates may vary over space or time, and so it is often crucial to collect these each time an acoustic survey takes place. Indeed, *A. lightfooti* call densities were not converted into *A. lightfooti* calling male densities in Section 3.5.3 using the call-rate data from Section 3.5.2 as these were collected on a single occasion—they may not be representative of the entire calling season across all sites. Whether the observed spatial and temporal effects on call density (Figure 3.5) were due to variation in calling male density, in call rates, or a combination of the two, therefore remains an open question.

One particular concern lies in the estimation of μ_r. Acoustic surveys are beneficial in situations where the focal species is visually cryptic, but vocal; detection of individuals' calls becomes the most efficient means of detecting the individuals themselves. However, obtaining a simple random sample of animals from which to collect the cue-rate data then becomes problematic if there is any between-individual heterogeneity in cue rates: those that call more often are likely to be overrepresented in the sample by virtue of being more easily detectable (Buckland et al., 2001, p. 192). This induces positive bias in the estimator $\widehat{\mu}_r$, thus negative bias in $\widehat{D}_a$.

3.6.2 Model selection

A boon of ML estimation is access to a wide variety of likelihood-based inference machinery that can be applied following parameter estimation. Examples include likelihood-ratio tests and likelihood-based information criteria, such as AIC and Bayesian information criterion (BIC)—both of which aid the decision of selecting the model that is 'best' supported by the data at hand, or alternatively allowing the incorporation of model uncertainty into parameter estimates via model averaging (Burnham & Anderson, 2002).

When the objective function being maximised for parameter estimation is not the likelihood for the model being fitted, access to these tools is lost. For example, the approach used here maximises a likelihood for a model that falsely assumes independence between capture histories generated by calls from the same individual. In such a case, the likelihood assumes the data are more informative than they truly are—thus likelihood-based CIs are typically

too narrow (see the coverage of the naïve method in Table 3.2) and information-theoretic model selection approaches tend to underpenalise additional parameters, resulting in the selection of overfitted models.

The lack of any appropriate model-selection tools is a disadvantage of the cue-based SECR methods presented here. For example, the choice of the functional form to use for the detection function can have a large enough impact on the parameters of interest for this decision to require theoretical, data-driven justification. Simulation-based approaches, however, offer an avenue to reclaim this lost ground.

Simulation-based likelihood-ratio tests

A likelihood-ratio-type test, for example, can be carried out by fitting both the null and alternative models to data generated from the former. For each set of generated data, a likelihood-ratio-type test statistic can be calculated (i.e., twice the difference in the log of the maximised objective functions), providing this test statistic's sampling distribution under the null hypothesis. The observed statistic from the real data can then be compared to this to conduct the hypothesis test. Such an approach is common when the likelihood-ratio test statistic is not asymptotically chi-squared; for example, see McLachlan (1987).

Simulation-based information criteria

The Kullback-Leibler discrepancy provides a measure of separation between two models. AIC aims to provide an estimate of a model's expected Kullback-Leibler discrepancy from the 'true' model; thus, the lower a fitted model's AIC, the closer it is likely to approximate the process from which the data were generated. Akaike (1973) noted that $-2\log(L(\widehat{\boldsymbol{\theta}};\boldsymbol{x}))$ (where $\boldsymbol{x}$ are the observed data, and $\widehat{\boldsymbol{\theta}}$ contains estimates of the k model parameters) provides an estimate of this expectation with bias that is asymptotically given by $-2k$. Therefore, $\text{AIC} = -2\log(L(\widehat{\boldsymbol{\theta}};\boldsymbol{x})) + 2k$ provides an asymptotically unbiased estimate.

One considerable drawback of AIC is the asymptotic justification of this bias correction: when the sample size is small and k is comparatively large, the 'true' bias can be substantially larger in magnitude than $-2k$ (Hurvich & Tsai, 1989). AIC then underpenalises additional parameters and tends to favour overcomplicated models—a similarity to what is noted above regarding the use of AIC with the objective function maximised for the cue-based SECR approaches presented in this chapter. Various bias-estimation and correction methods have been suggested (the most-enduring giving rise to AICc; Hurvich & Tsai, 1989). In particular, Cavanaugh and Shumway (1997) proposed AICb, which utilises a bootstrap procedure to estimate the bias rather than relying on the asymptotically justified

$-2k$. If a total of n_b bootstrap resamples are generated, and $\widehat{\boldsymbol{\theta}}_i^*$ contains the parameter estimates from the ith such resample, then

$$\text{AICb} = -2\,\log(L(\widehat{\boldsymbol{\theta}};\boldsymbol{x})) + 2\left[\frac{1}{n_b}\sum_{i=1}^{n_b} -2\,\log\left(\frac{L(\widehat{\boldsymbol{\theta}}_i^*;\boldsymbol{x})}{L(\widehat{\boldsymbol{\theta}};\boldsymbol{x})}\right)\right].$$

The first ('goodness-of-fit') term remains unchanged from standard AIC. Note that the numerator in the summand of the second ('bias-correction') term is the likelihood evaluated at the parameter estimates obtained from the resampled data, but calculated using the *observed* data—this will therefore necessarily be less than the denominator (the maximised likelihood), and so the bias correction is always positive.

One can therefore view this term as calibrating how large a difference in likelihoods can be expected between that evaluated at the ML estimate and that evaluated at another 'plausible' point in the parameter space—using a bootstrap is an alternative means to relying on the theoretical properties of a likelihood ratio to inform 'plausibility'. For example, consider a situation where parameter estimates from bootstrap resamples routinely return markedly lower likelihoods (when evaluated with the observed data) than the maximised likelihood. This would suggest that sets of parameters far from the ML estimate are more plausible than standard likelihood theory would otherwise suggest (as is the case with the cue-based SECR methods suggested here), and thus a larger penalty should be applied for additional parameters. This is precisely what happens with the AICb penalty term—large discrepancies in likelihoods result in a large sum within the bias-correction term, and therefore a larger bias correction than what standard AIC would otherwise provide.

Thus, in the absence of a true likelihood, AICb appears a promising alternative to standard information-theoretic model selection criteria for cue-based SECR methods. It can make use of the parameter estimates obtained during the simulation method (Section 3.2), and so—once this has been performed—calculation of AICb has a low marginal computational cost. Before being hailed as a solution to the model selection problem faced here, however, it is necessary to (i) further investigate the theory underpinning AICb in order to ascertain whether or not it is capable of appropriately estimating bias in Kullback-Leibler discrepancy in situations where the maximised objective function is not the model's true likelihood, and, if so, (ii) perform a simulation study to verify its use. Calculation of AICb is therefore not yet implemented in the `admbsecr` package in order to avoid its potential misuse.

3.6.3 Implementation in `admbsecr`

The software implementation of cue-based SECR methods in the `admbsecr` package aims to maximise their impact by facilitating the application to acoustic detection data. Care has been taken to automate each step—from data processing, through the analysis, and to the presentation of results. A testament to this is that, for the application described in Section 3.5.2, the entirety of this (relatively complicated) process can be achieved with little more than the three simple lines of code displayed therein. While some data processing had been undertaken in advance to generate the `lightfooti` data object, it is simple to reconstruct from the PAMGuard output data alone:

1. Call `convert.pamguard()` to ascertain which detections are of the same cue and generate the `capt` component in the correct format (as seen in Section 3.5.2). If the acoustic data were not processed with PAMGuard, then these steps can be carried out via calls to `make.acoustic.captures()` and `create.capt()` (see the `admbsecr` package documentation).

2. Create the `traps` component by specifying the coordinates of the detectors.

3. Generate the `mask` component via a call to `create.mask()`.

Creation of the components `cutoff` and `freqs` is trivial—these constitute a scalar and a vector, respectively.

Granted, the methods described here are computationally demanding, particularly the simulation procedure—the simplicity of the code does not imply that high-performance computing resources are not required, and it is likely that they are for larger data sets. Efforts have been made to minimise this computational cost in various says (see Section 2.3); for example, recall that the `boot.admbsecr()` function supports parallel processing. One caveat here, however, is that even a single process appears relatively memory expensive, and attempting to run too many in parallel may exhaust the system's available RAM.

3.6.4 Applications to *A. lightfooti* survey data

Alternative, previously used monitoring procedures for the genus *Arthroleptella* involve trained practitioners listening to an assemblage and subjectively placing abundance into a category (see Dorcas et al., 2009). The assemblage surveyed at Site 1 in Section 3.5.2 was assessed using this method, with the estimate falling into the highest category: 'more than 100 individuals'. This is consistent with the estimated *A. lightfooti* calling male density obtained here (366.08 individuals per hectare; Table 3.3) given the size of the seepage that

these frogs inhabit. The method used here is clearly preferable on two counts: it results in an estimate that (i) is quantitative and (ii) has a measure of uncertainty.

The application of acoustic SECR was particularly suited to the situation described in Section 3.5.3; in this case, *call* density, D, was of inherent interest in itself—the lack of available cue-rate data did not impact the investigation. Call density was estimated to peak in late July (Figure 3.5), and little support was found for models that included effects due to rainfall and temperature (Table 3.4).

Although no variance estimates were available for individual call-density estimates as a result of a lack of call-rate data, a measure of uncertainty was obtainable from the linear models fitted by generalised least squares due to the repeated sampling at the three sites over time—this provided estimates of error variance, and allowed calculation of the CIs for the expected call density (as a function of location and date) shown in Figure 3.5. However, it is important to note that the source of variance estimated here is different from that estimated by, say, a standard error for $\widehat{D}$ obtained from a single SECR survey. Regarding the estimation of spatial and temporal trends in call density, one can consider the estimate $\widehat{D}$ from a single survey as being subject to two sources of error: (i) the difference between the true call density and the expected call density as a function of date and location, and (ii) the difference between this true density and the estimate $\widehat{D}$ (i.e., due to sampling variation). A standard error for $\widehat{D}$ (for example, obtained using the simulation method described in Section 3.2) estimates only the variance of the latter, while the error variance estimated in the fitted linear model is a sum of the two. Indeed, the linear model confounds the two, and they cannot be estimated separately.

This is partly a consequence of taking a two-stage modelling approach, whereby estimates obtained from the SECR models were then used as the response for the linear model. Arguably, it would have been conceptually neater and more informative to fit a single model that simultaneously estimated (i) both spatial and temporal trends (possibly with other covariates) in call density, and (ii) the call detection function, along with the other SECR model parameters. Additionally, in such a case, there is potential for identifiability (and therefore estimation) of both variance components mentioned above. However, the fitting of such a model is far from straightforward, as (i) a full likelihood has not yet been specified for the SECR model component, and so it is unclear how a general estimation method can be constructed for all parameters under one model; and (ii) given the abundance of data, parameter estimation for such a model is unlikely to be computationally tractable—fitting a larger number of smaller models and combining them under a two-stage approach is far more efficient, and still required substantial computation.

3.6.5 Monitoring anuran populations

Traditional monitoring approaches can be laborious and time consuming; they often involve the physical capture of individuals, and rely on returning to the survey location on multiple occasions in order to obtain a single abundance or density estimate. This contrasts markedly with the surveys conducted to collect the data analysed in Section 3.5. These lasted just forty minutes each, and a single survey was sufficient to estimate call density (although the additional collection of cue-rate data is required to estimate animal abundance and density). This is a particular advantage if the focal species is in a remote location, as travel costs are substantially reduced.

Acoustic SECR methods thus have the potential to revolutionise the monitoring of anuran populations. In some cases, obtaining an equivalently precise abundance estimate with traditional monitoring approaches is likely to be associated with a fieldwork cost that is orders of magnitude larger—this is likely the case when vocalisations are easily heard, but individuals are not straightforward to detect by other means (e.g., by virtue of being visually cryptic). Moreover, acoustic SECR offers major advantages over the existing acoustic MCS and ARS methodologies. In particular, SECR offers *absolute* estimates of abundance and density—rather than *relative* changes (e.g., across time or space), or indices thereof. Additionally, SECR explicitly estimates detectability, thus accounting for variation in ESA across multiple surveys. Section 3.5.3—specifically Figure 3.6—has shown that detectability can vary substantially across sites (despite their geographical proximity), and it is likely that there is also variation over time. This raises serious questions about the appropriateness of MCSs and ARSs—neither make any effort to account for ESA, thus confounding density with detectability.

3.6.6 Concluding remarks

While this chapter presents a general method for animal density estimation from an array of acoustic detectors, the methods are underpinned by the notion that maximising a falsely specified likelihood is appropriate. A proof showing that the call density estimator, $\widehat{D}$, has desirable properties has not been presented, although the simulation studies in Section 3.4 provide empirical evidence of asymptotic unbiasedness and normality under the conditions tested.

The specification of a proper likelihood would represent a notable breakthrough, however without information about the identity of the individual associated with each detected cue this would likely be intractable. Nevertheless, the data hold information about animal identities that is not currently being utilised: the observed capture histories and auxiliary

data associated with each cue give rise to estimated locations. The more similar cues' location estimates are, the more likely it is that they were emitted by the same individual; thus there is potential for an estimation approach that directly estimates animal density without the collection of independent cue-rate data. Indeed, a means of estimating animal density from SECR survey data without any identity information is available (Chandler & Royle, 2012), however it is unsuitable for small arrays of detectors that detect a large number of individuals (as is typically the case with acoustic survey data). As an alternative, methods for animal density estimation from (non-SECR) surveys of unidentifiable individuals are developed in Chapter 6, and there is potential to generalise these to acoustic SECR scenarios.

There is also scope for the development of statistical methodology for surveys that *do* collect animal identity information from cues (see Section 7.1); the base method of Efford, Dawson, and Borchers (2009) does not incorporate this information and simply estimates cue density. Of course, identities could be discarded and the methods presented in this chapter could be employed, but these come with the requirement of independent cue-rate information. One approach that foregoes this restriction is presented in Chapter 4.

Chapter 4

First-cue SECR methods

4.1 Introduction

Analysing cue-based SECR capture histories (e.g., Table 1.1a) can be problematic as (i) only density estimates of cues (measured in cues per unit time per unit area), not individuals, are directly available from SECR survey data alone; and (ii) capture histories cannot be considered statistically independent if they correspond to cues produced by the same individual with unknown location. The methodology presented in Chapter 3 avoids these issues using independently collected cue rate data, allowing (i) cue density to be converted into animal density using the mean population cue-production rate, and (ii) the use of a simulation approach to account for intercue dependence. The disadvantages here are that individual identities associated with cues (if observed) are ignored, possibly decreasing estimator precision; collecting the independent cue-rate data can vastly increase the amount of field effort required; and if these data are not collected at the same place and time as the acoustic survey, then estimates may be biased due to spatial or temporal variation in cue rates.

If animals can be identified from their cues, then it is possible to construct individual-based capture histories (e.g. Table 1.1b), whereby $\omega_{ij} = 1$ if at least one cue produced by the ith individual was detected by the jth detector and $\omega_{ij} = 0$ otherwise. In such cases, standard, well-established SECR approaches for *binary proximity detectors* (those that detect the presence of an individual without physically restraining it, allowing future detection at other detectors) can be used to estimate animal detectability and density (Efford, Dawson, & Borchers, 2009). Alternatively, capture histories can be constructed so that ω_{ij} gives the number of detections made by the jth detector of the ith individual (and so an entry of 0 still indicates nondetection); Efford, Borchers, and Byrom (2009)

derived an appropriate likelihood for surveys on which these *count proximity detectors* are deployed, maximisation of which again results in suitable animal density and detectability estimators. However, use of these methods for the analysis of individual-based capture histories for both binary and count proximity detectors often precludes the incorporation of auxiliary information, as these data are typically cue based[1]—no estimation methods yet exist for situations where the sampling unit of the capture histories and the auxiliary data do not match. This is a notable drawback, as auxiliary information can make appreciable contributions to density estimator precision (Chapter 2, Figure 2.5, Table 2.1).

For cue-based models, $g(d;\boldsymbol{\gamma},c)$[2] returns the probability of detecting a *cue* at a detector situated distance d from the its location, and $p(\boldsymbol{x};\boldsymbol{\gamma},c)$ returns the probability of a *cue* emitted at $\boldsymbol{x}$ being detected by at least one detector, respectively. For individual-based models these return the probability of detecting an *individual* at a detector situated distance d from its location, and the probability of an *individual* located at $\boldsymbol{x}$ being detected by at least one detector. Thus, to estimate *animal* density it must be possible to estimate the probability of detecting an individual given its location, and this is available directly from $p(\boldsymbol{x};\boldsymbol{\gamma})$ if individuals are the unit of detection. However, this is not the case with standard SECR methods that analyse cue-based capture histories. In this case, the probability of detecting an individual is the probability of detecting at least one of its cues, and this clearly depends on how many cues it has produced—the more cues, the more chances it has to be detected. The probability of an individual's detection is therefore greater than or equal to the probability of a cue's detection. The number of cues each individual has made remains unobserved, as it is not known when a cue is produced but not detected. The probability of detecting individuals is not straightforward to calculate in this case.

It is desirable to combine advantages of both cue- and individual-based approaches, giving rise to an animal density estimator that does not require the collection of independent cue-rate data, and is also able to incorporate auxiliary information at the cue level. While this appears far from straightforward when individuals are not identifiable from their cues (Section 3.6.6), when they are there is potential for simultaneous estimation of both the cue-production and cue-detection processes directly from the capture history and auxiliary information data. There is scope for further methodological development in this regard; see Section 7.1.

The method used by Dawson and Efford (2009) for analysis of *S. aurocapilla* acoustic

[1]The *N. annamensis* data analysed in Chapter 2 is an exception to this: although observers could potentially make multiple detections of the same gibbon group, they only estimated a single group-based bearing rather than multiple cue-based bearings. The methods of Borchers et al. (2015) allow direct inference when both capture histories and auxiliary data are collected at the individual (rather than the cue) level.

[2]The constant c only appears here for models incorporating auxiliary signal strength data.

survey data offers a simpler alternative, and involves a new protocol for including cue detections' capture histories for analysis: (i) for each individual identified by a detector array, all cues other than the first detected are ignored, and (ii) for all first cues detected, only those detections with measured signal strengths above a threshold value, c are *included* for analysis. That is, $\omega_{ij} = 1$ if the received signal strength at the jth detector of the first detected cue produced by the ith detected individual exceeds c, and $\omega_{ij} = 0$ otherwise. The threshold c is set high enough so that any received signal exceeding c is distinguishable from any background noise and is therefore detected with certainty (although many cues may still be detected at strengths below c). Note the subtle difference here between an individual being 'detected' and 'included': an individual is detected if it is identified from a cue received by one or more detectors at any time during the survey, but only included if a received signal strength of the first cue that was detected exceeds c at one or more detectors. Individuals that are detected but not included therefore do not contribute any data to the analysis (even if a subsequent cue is detected at a strength in excess of c).

The idea behind including only those cues received at a strength in excess of c that were the first detected from an individual is to give each animal a single opportunity to produce a cue eligible for analysis (similar to conducting an instantaneous 'snapshot' survey, as recommended by Buckland, 2006). Otherwise, given a long enough survey, even those individuals situated quite some distance from the microphone array may eventually have a cue received at a high strength by virtue of producing a large number of calls. If an individual's probability of inclusion increases with the more calls it produces, then this must somehow be accounted for in the model—this is difficult given that it is not known how many cues each individual produces (although see Section 7.1).

Thus, the inclusion protocol of Dawson and Efford (2009) is attractive as (i) included capture history entries correspond to a particular cue, allowing the straightforward incorporation of auxiliary data; and (ii) each individual can only contribute a single capture history for analysis, allowing for an assumption of independence between included capture histories. Moreover, associating just a single cue with each individual avoids the requirement of considering the cue-production process—there is then no need to either collect independent cue-rate data or estimate cue-production parameters from the capture histories alone.

As above, an animal density estimator based on these first-cue data is appropriate as long as $p(\boldsymbol{x}; \boldsymbol{\gamma}, c)$ provides the probability of an *individual* located at $\boldsymbol{x}$ being included, thus allowing for the estimation of the density of individuals (detected or otherwise) that have not been included. However, Dawson and Efford (2009) simply applied the standard individual-based models of Efford, Dawson, and Borchers (2009) to their included data.

The capture histories included for analysis correspond to particular cues, and so the spatial range of detections within each of these provides information about the spatial detectability of cues, rather than individuals. The estimated detection function, $g(d; \widehat{\boldsymbol{\gamma}}, c)$, therefore also corresponds to cues, rather than individuals—that is, it returns the probability of a cue being detected by a detector situated a distance of x from its location. Therefore, given the DPS retaining its usual form of

$$p(\boldsymbol{x}; \boldsymbol{\gamma}, c) = 1 - \prod_{j=1}^{m} [1 - g(d_j(\boldsymbol{x}); \boldsymbol{\gamma}, c)], \tag{4.1}$$

this also returned the probability of detection of a *cue*; thus—as it was claimed that the parameter D referred to animal density—this application came with the implicit assumption that the probability of inclusion of a cue was equivalent to the probability of inclusion for an individual. As discussed above, the latter is necessarily greater than or equal to the former.

Consider the hypothetical (and unrealistic) situation in which all cues emitted within the survey region, $\mathcal{A}$, are detected. In this case, using the inclusion protocol of Dawson and Efford (2009), all cues other than the first produced by each individual are ignored, as any subsequent cues will not be the first to be detected. Thus, an individual's probability of inclusion is equivalent to the probability of their first cue being included. This probability is given by $p(\boldsymbol{x}; \boldsymbol{\gamma})$ when capture histories correspond to particular cues, and so the approach is appropriate.

However, this is not the case in general, as not all cues are detected. It is not known whether or not an individual's first detected cue was the first they produced. An animal located at $\boldsymbol{x}$ may produce a number of cues prior to detection, each of which had probability $p(\boldsymbol{x}; \boldsymbol{\gamma}, c)$ of exceeding strength c at one or more detectors and therefore being included. Thus, the probability of an individual being included is necessarily greater than the probability of a particular cue being included—and this holds whenever it is possible for subsequent cues to be included. This may cause the animal density estimator to be positively biased, as individuals' inclusion probabilities are not being appropriately calculated.

This chapter addresses the questions posed by acoustic SECR models that make use of capture histories comprising only animals' first detected cues. In some cases the consequences of the assumption violation discussed above are negligible, and diagnostic measures that can be used to ascertain whether or not this is the case are presented. More importantly, in Section 4.2, new SECR methodology is proposed, capable of appropriately modelling the process by which individuals' capture histories are included for analysis. The

same inclusion protocol of Dawson and Efford (2009) is employed; however, the probability of inclusion is calculated differently, to account for the fact that subsequent cues, as well as first cues, may be included for analysis. This results in a suitable animal density estimator for first-cue SECR models that allows straightforward incorporation of auxiliary data and does not require any cue-rate information. The implementation of the methodology developed in the `admbsecr` package, along with associated computational challenges, is discussed in Section 4.3. A simulation study is presented in Section 4.4, showing that the new estimator has favourable properties, and that using the method of Dawson and Efford (2009) can result in density estimators that are both imprecise and severely biased. The new methods result in models that are no more complicated than that of Dawson and Efford (2009) in terms of the number of parameters estimated—in fact, the density estimators from each converge when cue detection probability is the same as individual detection probability, as implicitly assumed by Dawson and Efford (2009). The new estimator will therefore always perform at least as well as the existing method at virtually no additional inferential cost. In Section 4.5, this is illustrated through reanalysis of the *S. aurocapilla* data used by Dawson and Efford (2009).

4.2 Methodology

The likelihood maximised by Dawson and Efford (2009) is equivalent to that presented in Chapter 2, incorporating signal strength information as auxiliary data. In this section signal strength information is again assumed to be the only auxiliary data available, however incorporation of further auxiliary data types is straightforward via extensions equivalent to those presented in Chapter 2.

This method is extended to provide an appropriate animal density estimator for first-cue SECR data. Chiefly, it is necessary to disentangle the probabilities of inclusion for individuals and cues, as they are not generally equivalent. This is achieved in Section 4.2.1. Estimation of $\boldsymbol{\theta}$ is then detailed in Section 4.2.2.

4.2.1 Individual probability of inclusion

First, it is crucial to formally define the detection and inclusion of cues and individuals in the context of the methodology described in this chapter. Here, two thresholds are set: a lower threshold, c_l, and an upper threshold, c_u. Let $\boldsymbol{c} = (c_l, c_u)$. The probability that a cue is received with a strength exceeding some general level c by at least one detector remains

as per Equation (4.1), but is now redefined as $p_c(\boldsymbol{x};\boldsymbol{\gamma},c)$; that is,

$$p_c(\boldsymbol{x};\boldsymbol{\gamma},c) = 1 - \prod_{j=1}^{m}[1 - g(d_j(\boldsymbol{x});\boldsymbol{\gamma},c)]. \tag{4.2}$$

While various signal strength detection functions are described in Section 2.2.3, in this chapter that with an identity link function is assumed for simplicity. Recall that this gives

$$g(x;\boldsymbol{\gamma},c) = 1 - \Phi\left(\frac{c - (\beta_0 - \beta_1 x)}{\sigma_s}\right). \tag{4.3}$$

A cue produced at the location $\boldsymbol{x}$ is considered to be detected by the jth detector if it is received at a strength above the lower threshold, c_l, and this occurs with probability $g(d_j(\boldsymbol{x});\boldsymbol{\theta},c_l)$. It is detected in general if at least one detector receives the signal at a strength exceeding c_l, and so this occurs with probability $p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)$. Cues are included only when their received signal strength is above the threshold c_u instead of c_l. Cue inclusion probabilities are therefore $g(d_j(\boldsymbol{x});\boldsymbol{\theta},c_u)$ and $p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)$ at detectors and by the whole array, respectively. The lower threshold may be considered a lower limit on the ability of the detectors to record cues strong enough for them to be identified as such with certainty. For example, with acoustic surveys c_l might represent a level above which received acoustic signals are loud enough to be discriminated from background noise, and also to allow recognition of individuals. It is always known when a cue is detected but not included; however, it is not necessarily known when a cue is emitted but neither detected nor included.

An individual that is detected is also included if its first detected cue is also an included cue (i.e., the received strength of its first detected cue not only exceeds c_l, but also c_u) at one or more detectors. The likelihood described in Chapters 1 and 2 is applied to the capture histories and signal strengths associated with included cues from the included individuals (although $\boldsymbol{c}$ now appears as an argument to many functions). The probability of an individual being included is still defined as the function $p(\cdot)$, but no longer takes the same form as Equation (4.1). Recall that it appears within the likelihood in both the PMF of the number of included individuals (Equation (1.7)) via its expectation,

$$\mu = D\int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{c})\,\mathrm{d}\boldsymbol{x},$$

and the marginal PDF of the location of the ith included individual (Equation (1.4)),

$$f(\boldsymbol{x}_i;\boldsymbol{\gamma},\boldsymbol{c}) = \frac{p(\boldsymbol{x}_i;\boldsymbol{\gamma},\boldsymbol{c})}{\int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{c})\,\mathrm{d}\boldsymbol{x}}.$$

The DPS for first-cue SECR models, $p(\boldsymbol{x};\, \boldsymbol{\gamma},\, c)$, is derived below.

An individual is included if its first detected cue is included. Recall that this occurs when either (i) its first produced cue has a received signal strength above c_u at one or more detectors, or (ii) one of its subsequent cues has a received signal strength above c_u at one or more detectors, and all previous cues did not have a received signal strength above c_l at any detector.

Let the probability of an individual located at $\boldsymbol{x}$ being included via its first cue (i.e., due to (i), above) be $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, and the probability of being included via a subsequent cue (i.e., due to (ii), above) be $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$. The probability of an individual at $\boldsymbol{x}$ being included is therefore given by

$$p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) + p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}). \tag{4.4}$$

Any cue, including the first produced by an individual during a survey, is only detectable once, and so $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \equiv p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)$; that is, the probability of any given cue being included. The method of Dawson and Efford (2009), on the other hand, assumes $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \equiv p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)$ and, therefore, that $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = 0$. That is, if it is assumed that individuals are only detectable once, then the probability of including an animal located at $\boldsymbol{x}$ is equivalent to the probability of including a cue produced at $\boldsymbol{x}$. If individuals are detectable more than once (e.g., by producing more than one cue), then this no longer holds. The function $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ is derived for such a case below.

The probability that an individual located at $\boldsymbol{x}$ produces k cues that remain undetected before its $(k+1)$th cue is subsequently included (thus resulting in the animal's inclusion via a subsequent cue) is given by $p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]^k$. Once an individual is included, it is not known how many previous cues it produced prior to its detection and inclusion. Calculating $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ must therefore sum over all possible values for the number of undetected cues that were produced. Given a long enough survey, $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ can be approximated by assuming that each individual eventually produces a detected cue (in theory, this occurs with certainty given a long enough survey; see Section 4.6.2 for discussion on how the model is robust to departures from this). This gives

$$p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\sum_{k=1}^{\infty}\left[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)\right]^k. \tag{4.5}$$

This is a convergent geometric series. The sum of its first q terms is

$$p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\sum_{k=1}^{q}\left[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)\right]^k = p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\left(\frac{1-[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]^{q+1}}{1-[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]}-1\right)$$

$$= p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\left(\frac{1-[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]^{q+1}}{p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)}-1\right),$$

and therefore

$$\begin{aligned} p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) &= \lim_{q\to\infty} p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\sum_{k=1}^{q}[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]^k \\ &= \lim_{q\to\infty} p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\left(\frac{1-[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]^{q+1}}{p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)}-1\right). \end{aligned}$$

As $q\to\infty$ the term $[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]^{q+1}\to 0$, and so this gives

$$\begin{aligned} p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) &= p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)\left(\frac{1}{p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)}-1\right) \\ &= \frac{p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]}{p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)}. \end{aligned} \tag{4.6}$$

An alternative derivation is also available. Omitting subscript i for clarity, recall that y_j^* gives the received signal strength of a particular cue at the jth detector, regardless of whether or not it made a detection (Section 2.2.3); therefore, y_j^* is only sometimes observed below c_l, but if a received signal strength exceeds this level then y_j^* is observed with certainty. Let $y' = \max(\boldsymbol{y}^*)$, and therefore $p_c(\boldsymbol{x};\boldsymbol{\theta},c) \equiv \Pr(y' > c|\boldsymbol{x};\boldsymbol{\theta})$; that is, the probability of at least one detector receiving a signal stronger than some threshold c is equivalent to the probability of the strongest received signal exceeding c. Recall that a subsequent cue from an individual is included if (i) the received signal strength of its first cue is not received above c_l at any detector, and so $y' < c_l$; and (ii) the first cue it produces that is received by any detector with a signal strength above c_l is also received above c_u by one or more detector. Thus $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ can be obtained as a product of the probabilities associated with the events outlined in (i) and (ii), above:

$$\begin{aligned} p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) &= \Pr(y' < c_l;\boldsymbol{x},\boldsymbol{\theta})\Pr(y' > c_u|y' > c_l;\boldsymbol{x},\boldsymbol{\theta}) \\ &= [1-\Pr(y' > c_l;\boldsymbol{x},\boldsymbol{\theta})]\frac{\Pr(y' > c_u \cap y' > c_l;\boldsymbol{x},\boldsymbol{\theta})}{\Pr(y' > c_l;\boldsymbol{x},\boldsymbol{\theta})} \\ &= [1-\Pr(y' > c_l;\boldsymbol{x},\boldsymbol{\theta})]\frac{\Pr(y' > c_u;\boldsymbol{x},\boldsymbol{\theta})}{\Pr(y' > c_l;\boldsymbol{x},\boldsymbol{\theta})} \\ &= \frac{p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]}{p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)}, \end{aligned}$$

as per Equation (4.6).

From Equations (4.4) and (4.6), the probability of including an individual located at $\boldsymbol{x}$ is

$$p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = p_c(\boldsymbol{x};\boldsymbol{\theta},c_u) + \frac{p_c(\boldsymbol{x};\boldsymbol{\theta},c_u)[1-p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)]}{p_c(\boldsymbol{x};\boldsymbol{\theta},c_l)}. \tag{4.7}$$

Employing the method of Dawson and Efford (2009) may result in approximately unbiased inference even when some cues remain undetected, so long as $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \approx p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, or, equivalently, $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \approx 0$ for all $\boldsymbol{x} \in \mathcal{A}$. To ascertain whether or not this is the case it is useful to plot $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, and $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ against $d_\cdot(\boldsymbol{x})$, the Euclidean distance between $\boldsymbol{x}$ and the detector centroid, for a discrete grid of points spanning $\mathcal{A}$ (e.g., the habitat mask points). See Figure 4.2 for examples taken from the simulation study and the application—presented in Sections 4.4 and 4.5, respectively.

4.2.2 Estimation

Estimation of $\boldsymbol{\theta}$ is achieved via maximisation of the likelihood described in Chapter 2 (Equation (2.1))—the only required change being in the calculation of $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ to Equation (4.7) from that shown in Equation (4.1).

Now that each individual only contributes a single capture history (and a single set of auxiliary information) for analysis, independence across included cues is potentially appropriate. If individuals can be considered independent the likelihood is a proper likelihood, and so—unlike with the use of methods described in Chapter 3—standard likelihood-based inference is possible. Variance estimates for the parameters of interest can therefore be calculated from the inverse of the observed Fisher information matrix. Alternatively, a parametric bootstrap can be used, for example, in situations where the parameters' sampling distributions may not be closely approximated by a normal distribution.

4.3 Implementation in `admbsecr`

Fitting a first-cue model using `admbsecr` is a straightforward extension to the process described in Section 2.3. All that is required is to specify a component named `lower.cutoff` in the list passed as the `ss.opts` argument to the `admbsecr()` function—this provides the lower threshold c_l. The `cutoff` component is then considered as the upper threshold, c_u. The inclusion of the `lower.cutoff` component results in the fitting of a first-cue model via the methodology described above.

A plot showing $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, and $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ against $d_\cdot(\boldsymbol{x})$ can be generated using the `mask.test()` function. The only argument required is a fitted model object.

Alternatively, `mask` and `traps` objects can be provided for a hypothetical survey, along with parameter values, the lower threshold, and the upper threshold (via the arguments `pars`, `cutoff`, and `lower.cutoff`, respectively).

4.3.1 Computation of subsequent cue inclusion probabilities

While this represents an extremely simple syntactical adjustment, it results in considerable changes to the estimation procedure carried out by the `admbsecr()` function. These are caused by challenges associated with the computation of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ (Equation (4.6)).

Consider computation of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ at locations near the edge of the survey region, far from the detector array. As distance from the detector array centroid, $d_\cdot(\boldsymbol{x})$, increases, $g(d_j(\boldsymbol{x});\boldsymbol{\theta},c) \to 0$ for all j and for any c (due to the signal strength detection function having a horizontal asymptote at 0). As a result, $\prod_{j=1}^{m}[1-g(d_j(\boldsymbol{x});\boldsymbol{\theta},c)] \to 1$, and therefore $p_c(\boldsymbol{x};\boldsymbol{\theta},c) \to 0$, for any c. This causes both $p_c(\boldsymbol{x};\boldsymbol{\theta},c_u) \to 0$ and $p_c(\boldsymbol{x};\boldsymbol{\theta},c_l) \to 0$.

These terms appear in the numerator and the denominator of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ (Equation (4.6)), respectively, causing two issues: (i) the behaviour of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ as $d_\cdot(\boldsymbol{x})$ increases is not clear, and (ii) its computation at locations further from the array result in the calculating of a quotient of two small values. The latter is problematic, as, given a quotient of two sufficiently small values, numerical underflow caused by the use of floating-point arithmetic results in inaccurate or incomputable probabilities—in the worst case, one or both of the probabilities are rounded to 0. Naïve approaches of computing $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ therefore fail.

The two issues outlined above can be overcome. First, it can be shown that $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \to 0$ as $d_\cdot(\boldsymbol{x}) \to \infty$ (i.e., the probability of inclusion decreases to 0 as the distance from the detector array increases; this result can be seen in Figure 4.2 and found in an appendix immediately following this chapter). Second, the computation of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ can be carried out with increased numerical stability using an alternative approach that partially relies on the above result.

In the calculation of $p_c(\boldsymbol{x};\boldsymbol{\theta},c)$, rather than subtracting the probability of all received signal strength being less than c from 1 (as per Equation (4.2)), it is necessary to enumerate all possible combinations of capture histories that result in at least one detector receiving a signal strength above c. The log of the probabilities associated with each combination can be calculated, then exponentiated and summed. That is, let

$$\mathcal{B} = \left\{ \boldsymbol{\omega} : \sum_{j=1}^{m} \omega_j \geq 1, \omega_j \in \{0,1\} \, \forall \, j \in \{1,\cdots,m\} \right\}, \tag{4.8}$$

and then computation of $p_c(\boldsymbol{x}; \boldsymbol{\theta}, c)$ can be achieved using

$$p_c(\boldsymbol{x}; \boldsymbol{\theta}, c) = \sum_{\boldsymbol{\omega} \in \mathcal{B}} \exp\left(\sum_{j=1}^{m} \log(f(\omega_j | \boldsymbol{x}; \boldsymbol{\theta}, c)) \right), \tag{4.9}$$

where $f(\omega_j | \boldsymbol{x}; \boldsymbol{\theta}, c)$ is the PMF

$$f(\omega_j | \boldsymbol{x}; \boldsymbol{\theta}, c) = \begin{cases} g(d_j(\boldsymbol{x}); \boldsymbol{\theta}, c) & \omega_j = 1; \\ 1 - g(d_j(\boldsymbol{x}); \boldsymbol{\theta}, c) & \omega_j = 0. \end{cases} \tag{4.10}$$

This involves the computation of the detection function—a probability from a standard normal CDF (Equation (4.3)). The log of this probability can be calculated with far greater stability than the probability itself in situations where its magnitude approaches machine precision. Moreover, the sum of the log-probabilities can be computed with more stability than the products of the probabilities themselves. These aid in the prevention of numerical underflow in the calculation of $p_c(\boldsymbol{x}; \boldsymbol{\theta}, c)$.

While this approach improves the accuracy in the calculation of the numerator and denominator of $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$, it does not eliminate the component of numerical instability that is induced by computing the quotient of two values with orders of magnitude nearing machine precision. Thus, $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$ is truncated (i.e., set to 0) when the denominator $p_c(\boldsymbol{x}; \boldsymbol{\theta}, c_l)$ is sufficiently small in magnitude. This is justified by the result mentioned above (i.e., that $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) \to 0$ as $d.(\boldsymbol{x}) \to \infty$; see the appendix following this chapter). Additionally, this appendix shows that if $c_l = c_u$ then $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) \to 1$ as $d.(\boldsymbol{x}) \to \infty$, rendering animal density inestimable, further motivating the requirement of lower and upper threshold values.

4.3.2 Estimation

This truncation causes the computed likelihood function to be nondifferentiable—as a result, derivatives calculated analytically via AD were found to be inaccurate (considerably so in some cases). This resulted in poor performance of an initial implementation of these methods that used ADMB for numerical optimisation of the log-likelihood. Instead, first-cue SECR methods are fitted in `admbsecr` using the implementation of the Nelder-Mead algorithm in the `nmk()` function of the `dfoptim` R package (Varadhan & Borchers, 2011). This provided more stability and accuracy in estimation than other implementations (e.g., the `optim()` function in the `stats` package). Calculation of the log-likelihood is carried out using compiled C++ code for increased efficiency.

Despite using a different optimiser to other methods implemented in the `admbsecr` package, the object returned maintains the same structure. All utility functions provided (e.g., `summary()` and `locations()`) can still be used with first-cue SECR model objects.

4.3.3 Data simulation

The `boot.admbsecr()` function can be used to calculate variance estimates via a parametric bootstrap. In order to do this it is necessary to simulate data under the fitted model. This might be done as follows:

1. Simulate animal locations within $\mathcal{A}$ from a Poisson point process with homogeneous intensity D
2. For each individual, continue to simulate received signal strengths from successive cues, conditional on its location, until at least one detector receives a signal strength above c_l
3. If the loudest received signal strength from a detection of this cue is greater than c_u, then this individual has been included, and this cue's capture history and auxiliary data are available for analysis; otherwise, this individual does not contribute data for analysis and is ignored

This naïve approach is again problematic: some locations within $\mathcal{A}$ may be too far from any detector to allow a received signal strength greater than c_l with a non-negligible probability. One must then set an upper limit on the number of cues each individual produces, declaring it to be excluded once this is reached. However, setting this too low results in some individuals' simulated detection probabilities, $p(\boldsymbol{x}_i; \widehat{\boldsymbol{\theta}}, \boldsymbol{c})$, being substantially smaller than they are under the estimated model; setting this too high results in a large computational cost associated with the simulation of large amounts of data for an individual that is not going to be included in any case. Such an approach is therefore at best inefficient, and at worst wildly misleading.

For the ith simulated individual, located at $\boldsymbol{x}_i$, an efficient, accurate alternative is available:

1. Select an element of $\mathcal{B}$ (Equation (4.8)) at random, weighted by the probabilities of selection given by $f(\boldsymbol{\omega}|\boldsymbol{x}_i; \boldsymbol{\theta}, \boldsymbol{c}) = \exp(\sum_{j=1}^{m} \log(f(\omega_j|\boldsymbol{x}_i; \boldsymbol{\theta}, c_l)))$—the summand in Equation (4.9) with the threshold set at c_l.
2. Simulate a received signal strength at all detectors at which a detection was made; that is, simulate y_{ij} if $\omega_{ij} = 1$. Recall that the received signal strength of a detected

cue is a random variable from a truncated normal distribution (with truncation set at c_l); its PDF is given by Equation (2.12).

3. If at least one generated signal strength is above c_u then the cue is included; if not, then this first detected cue has not been included, and the individual does not contribute any data for analysis.

Efficient simulation of random variables from the truncated normal distribution can be achieved using an accept-reject sampler (Geweke, 1991), even for parameter values that result in the support of the distribution being in an extreme tail of the corresponding untruncated normal PDF. Both the `sim.capt()` and `boot.admbsecr()` functions in the `admbsecr` package make use of the implementation of this algorithm found in the `truncnorm` R package (Trautmann, Steuer, Mersmann, & Bornkamp, 2014).

4.4 Simulation study

In all, 500 sets of capture histories and corresponding received signal strengths were simulated under the model described above for an array of sixteen detectors, arranged in a four-by-four grid with a spacing of 21 m between adjacent detectors. The simulation algorithm outlined in Section 4.3 (implemented in the `sim.capt()` function) was used to generate these data. Density was set at $D = 20$ animals per hectare, with $\beta_0 = 60$, $\beta_1 = 0.1$, $\sigma_s = 5$, $c_l = 52.5$, and $c_u = 55$ chosen specifically to illustrate a situation where the method of Dawson and Efford (2009) was hypothesised to perform particularly poorly; that is, with expected signal strengths decreasing slowly with cue propagation distance, and with similar upper and lower thresholds—thus there are large probabilities associated with many animals having their first cue undetected, but a subsequent cue included. Each simulated data set was analysed using the method presented in Section 4.2—using the `admbsecr` package—and the method presented by Dawson and Efford (2009)—using the `secr` package. The `admbsecr` package is also capable of fitting the latter model (providing equivalent results), but for the sake of comparison both the same method and software implemented by Dawson and Efford (2009) were used here.

Boxplots of all parameters estimated from the simulated data using the two approaches are shown in Figure 4.1. Estimator bias, precision, and mean-squared error (MSE) of both models for all parameters are shown in Table 4.1. For this given set of parameter values, thresholds, and detector layout the method presented in this chapter resulted in estimators with negligible bias, while that of Dawson and Efford (2009) resulted in very biased, imprecise estimators.

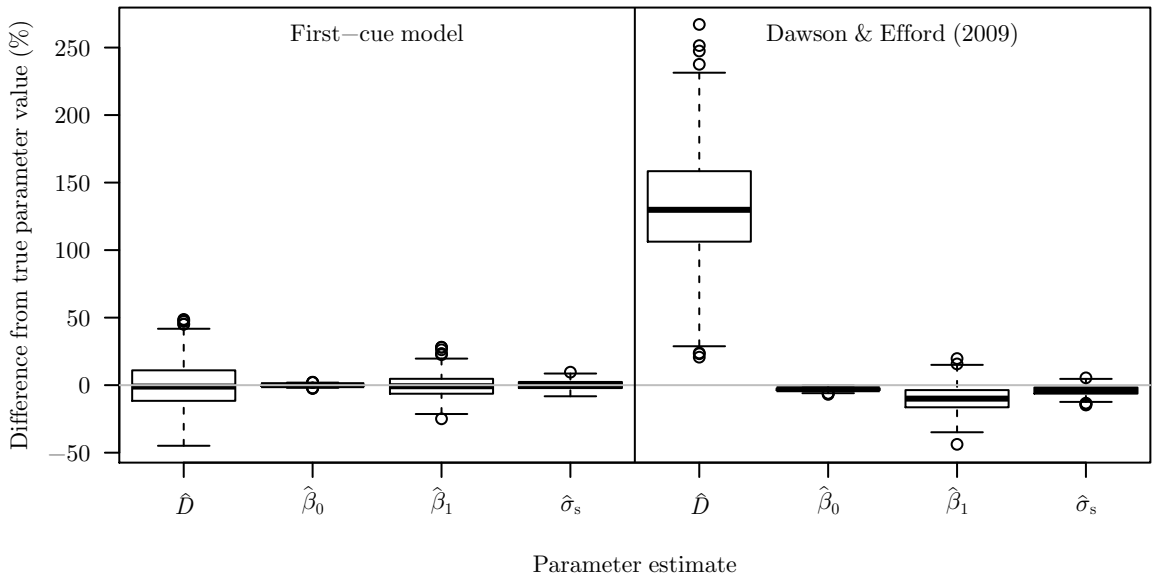


Figure 4.1 *Boxplots showing point estimates for all parameters using both models. These are given as percentage differences from the true parameter values from which the data were generated.*

Table 4.1 *Estimated biases, CVs, and RMSEs of estimators for the model parameters, all as percentages of their underlying values. These are presented for the first-cue models presented here (FC), and the method used by Dawson and Efford (2009; D&E)*

	Bias (%)		CV (%)		RMSE (%)	
Parameter	FC	D&E	FC	D&E	FC	D&E
D	−0.32	131.03	16.47	39.88	16.47	136.97
β_0	−0.05	−3.08	0.78	1.06	0.78	3.26
β_1	−0.35	−9.78	8.46	9.76	8.46	13.82
σ_s	0.17	−3.94	3.24	3.42	3.25	5.22

A plot of $d_\cdot(\boldsymbol{x})$ against $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, and $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ is shown in Figure 4.2(iv).

4.5 Application to *S. aurocapilla* survey data

Dawson and Efford (2009) recorded birdsongs of *S. aurocapilla* on a four-detector array in Patuxent Research Refuge near Laurel, Maryland, USA. The received signal strengths of the first songs that were detected from each individual are available in the `secr` R package (Efford, 2015) as an exported data object named `signalCH`. Their analysis made use of this package to estimate $\boldsymbol{\theta}$, although the same analysis can be carried out with the `admbsecr` package. In order to do this, first the data can be converted into objects suitable for use as

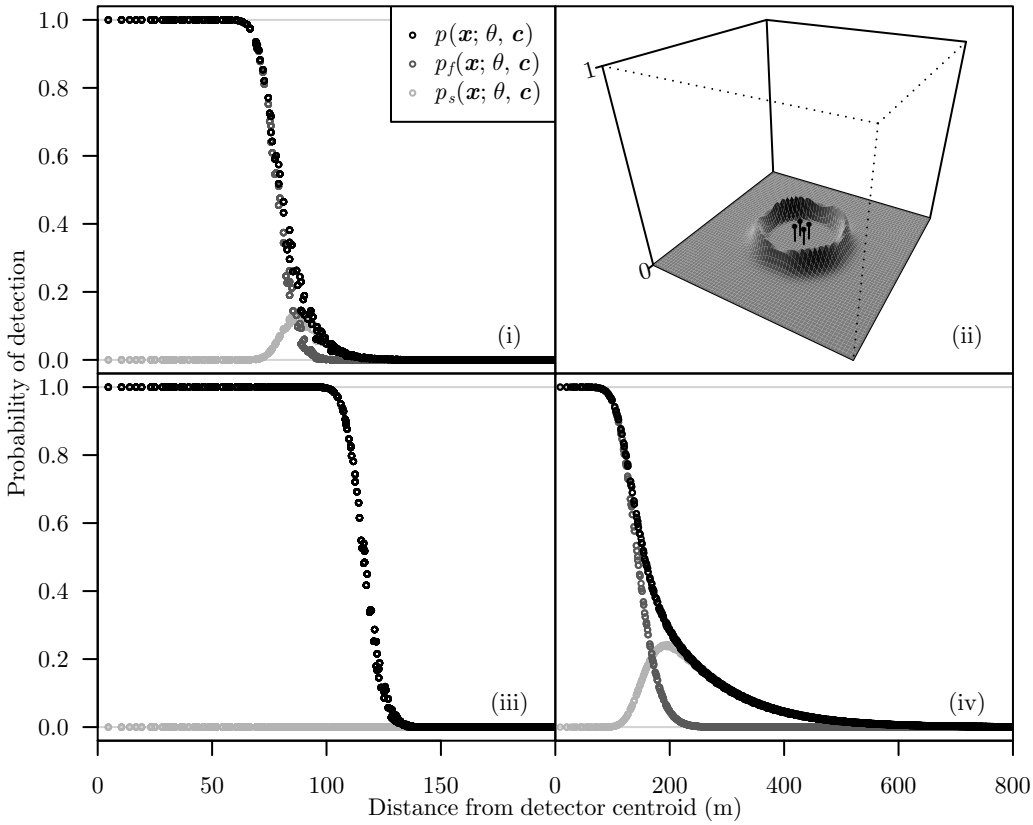


Figure 4.2 *Various inclusion probabilities across space for detector layouts from the real and simulated data. For Plots (i), (iii), and (iv), black points represent the probability of including an individual,* $p(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$*; dark grey points represent the probability of an individual's first cue being included,* $p_f(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$*; and light grey points represent the probability of an individual's subsequent cue being included,* $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$*, all calculated at a grid of discrete points spanning* $\mathcal{A}$*. The top row shows inclusion probabilities for parameter and threshold values for the detector array used to collect the* S. aurocapilla *data, with the hypothetical thresholds* $c_l = 60$ *and* $c_u = 62$*. Plot (i) shows the distance of points from the detector centroid against the inclusion probabilities. Plot (ii) shows* $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$ *across the domain* $\mathcal{A}$*, where pins show detector locations. It plots the same information as the light grey points in Plot (i), but over two dimensions rather than one. Plot (iii) shows the distance of points from the detector centroid against the inclusion probabilities for the detector layout, estimated parameter values and the threshold levels from the real data analysis—that is,* $c_l = 38.4$ *and* $c_u = 52.5$*. Plot (iv) shows the distance of points from the detector centroid against the inclusion probabilities for the detector layout, parameter values, and threshold levels from which data were generated for the simulation study. From Equation* (4.7)*,* $p(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) = p_f(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) + p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$*, and the method of Dawson and Efford (2009) assumes that* $p(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) = p_f(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$*. For the ovenbird data analysis, setting the lower threshold to just below the strength of the weakest received signal results in* $p(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) \approx p_f(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c})$*. The black points therefore overplot the dark grey points in Plot (iii), and the method of Dawson and Efford (2009) returns equivalent parameter estimates to the method presented in this chapter. At the hypothetical* S. aurocapilla *threshold levels (*$c_l = 60$ *and* $c_u = 62$*) and for those used to generate data for the simulation study,* $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) \not\approx 0$ *for some* $\boldsymbol{x}$*, in which case the method presented in this chapter models inclusion probabilities appropriately, while that of Dawson and Efford (2009) does not.*

arguments to the `admbsecr()` function:

```
capt <- convert.capt.to.admbsecr(signalCH)
traps <- traps(signalCH)
mask <- create.mask(traps, buffer = 200)
```

The model can then be fitted:

```
fit.DE <- admbsecr(capt = capt, traps = traps, mask = mask,
                   ss.opts = list(cutoff = 52.5))
```

The cutoff value set above is equivalent to that used by Dawson and Efford (2009).

Prior to fitting a first-call model (using the methodology set out in Section 4.2) it is necessary to set the lower and upper thresholds. Recall that c_l must be set at a strength such that all songs received by detectors above c_l are detected, and have a corresponding, observed value for y^*_{ij} (regardless of whether or not they are above c_u). The analysis of Dawson and Efford (2009) had the even stronger assumption that all songs within the survey region were detected so that every included song was a first song by an individual—this gives $p_s(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) = 0$, resulting in their calculation of $p(\boldsymbol{x}; \boldsymbol{\theta}, c)$ being appropriate—although this was not explicitly stated, and almost certainly violated. As models fitted by Dawson and Efford (2009) did not have any notion of a lower threshold, the value c_l should take with these data is unclear.

It would be suitable to set the lower threshold, c_l, above the highest level of background noise observed during the survey—approximately 60 dB. The upper threshold is then required to be set slightly larger; for example, $c_u = 62$. However, this requires that any song with a loudest received strength below this value be ignored (i.e., declared undetected due to the definition of a 'detection', as per Section 4.2), and so a subsequent song could plausibly generate data included for analysis by having a received signal strength above both c_l and c_u. Data of such subsequent songs are not available in the `secr` package, and have not been extracted from the recordings (M. G. Efford, personal communication, 15 June 2015), thus this analysis cannot be carried out here.

For the purposes of illustrating the use of the method presented in this chapter, c_l was instead set just below the weakest received strength. This is easily achieved in `admbsecr`:

```
fit.first_cue <- admbsecr(capt = capt, traps = traps, mask = mask,
                          ss.opts = list(lower.cutoff = 38.4,
                                         cutoff = 52.5))
```

However, this model likely violates the assumption that all songs above c_l are detected—Dawson and Efford (2009) reported fluctuating background noise that often exceeded this level, and this would have prevented detection of many received signals louder than c_l.

Setting c_l just below the weakest received signal strength resulted in point and variance parameter estimates that were virtually indistinguishable from those presented by Dawson and Efford (2009) for the corresponding model. A plot of $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, and $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ against $d.(\boldsymbol{x})$ (Figure 4.2(iii)) shows the reason for this: the subsequent-cue inclusion probability $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ is negligible for all points in $\mathcal{A}$, and so inclusion probabilities calculated using the method of Dawson and Efford (2009) are virtually equivalent to those calculated using Equation (4.7).

It is of interest to investigate if this result holds with threshold settings that do not violate model assumptions—for example using $c_l = 60$ and $c_u = 62$, as suggested above. A plot of $d.(\boldsymbol{x})$ against $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, $p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, and $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ with these hypothetical threshold settings—still using the same parameter values estimated from the *S. aurocapilla* analysis—is shown in Figure 4.2(i). Now, $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \not\approx 0$ for some $\boldsymbol{x} \in \mathcal{A}$ (Figure 4.2(ii) and (iii)), indicating that, in this case, the two methods no longer calculate near-equivalent inclusion probabilities.

4.6 Discussion

Using only the first detected cue from individuals detected by a detector array is an attractive option. At present it is the only SECR data-collection approach that generates animal (rather than cue) density estimates from cue-based capture histories without the requirement of collecting additional cue-rate data. It therefore has the flexibility to incorporate auxiliary information at the individual cue level. The methodology presented here is capable of generating animal density estimates that have negligible bias, and do not rely on unrealistic assumptions regarding the ability of a detector array to detect all cues produced within the survey region.

4.6.1 Method comparison

The shape of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ shown in Figures 4.2(ii) was expected, whereby $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \approx 0$ across areas of the domain $\mathcal{A}$ both close to the detector array and far away, but not necessarily in between. Individuals close to the detector array may have a very low probability of their first cue being received by all detectors below c_l due to this proximity (i.e., $p_c(\boldsymbol{x};\boldsymbol{\theta},c_l) \approx 1$ if $d.(\boldsymbol{x})$ is small), and so their first cue has a high probability of being detected. Any subsequent

cue is therefore ignored. Individuals far from the array may have a high probability of their first cue remaining undetected by all detectors; however, due to this large distance there is little chance that their first cue that is eventually detected exceeds not only c_l, but also c_u; indeed, the appendix at the end of this chapter shows that this probability tends to zero as the distance from the detector centroid increases. Thus, it may only be those individuals at moderate distances from the array that have a chance of having a subsequent cue included—there are non-negligible probabilities associated with both their first cue going undetected and their first subsequent cue that is detected also being included. The probability of this occurring decreases as the difference between the upper and lower thresholds increases, that is, $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \to 0$ as $c_u - c_l \to \infty$: as this difference increases it becomes increasingly unlikely that an individual could be far enough from the array for its first cue to remain undetected, yet close enough for the first subsequent cue that is detected to also be included.

Thus, with c_u fixed, $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \to 0$ as $c_l \to -\infty$. Recall that the likelihood maximised by Dawson and Efford (2009) to obtain parameter estimates is equivalent to that presented here when $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = p_f(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$; that is, when $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = 0$ for all $\boldsymbol{x} \in \mathcal{A}$. From Figure 4.2(iii), this appears to be the case with the real data analysis when c_l was set at a level below the weakest detected strength—the equivalence of parameter estimates between the two methods is therefore not surprising. The method of Dawson and Efford (2009) can thus be viewed as a special case of the model presented here. Their method assumes that all cues are detected, and so it is known that each cue included corresponds to the first cue that the individual emitted; this occurs under the framework presented in this chapter when $c_l \to -\infty$.

The generalisation presented here is important. In almost all cases it is unrealistic to assume that every cue made by individuals located within $\mathcal{A}$ will be detected by at least one detector in the array. This was not explicitly stated as an assumption of the approach used by Dawson and Efford (2009); however, it has been shown here that a violation of this assumption can result in large bias in the animal density estimator under some conditions (Section 4.4, Figure 4.1, Table 4.1). The method of Dawson and Efford (2009) can be robust to violations of this assumption as long as the difference between the two thresholds is sufficiently large (as per the above discussion), and the diagnostic plots generated by the `mask.test()` function in the `admbsecr` package (shown in Figure 4.2) are a useful tool to ascertain whether or not this is the case.

The estimator derived here is appropriate regardless of whether or not $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \approx 0$ for all $\boldsymbol{x} \in \mathcal{A}$, and so its use is recommended in place of that used by Dawson and Efford

(2009). When the latter is robust to the aforementioned assumption violation, using the first-cue model presented here will simply yield equivalent results—the only loss being a modest amount of computation time associated with the calculation of $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ across $\mathcal{A}$. When it is not, estimates from first-cue model are preferable (Figure 4.1, Table 4.1).

4.6.2 Survey length

Any survey of finite length may miss individuals because they do not produce a cue during the survey period, and hence it may only estimate only the density of individuals that have produced a cue rather than that of all cue-producing individuals. There are two ways around this problem: one may either (i) make the survey long enough to avoid this, or (ii) obtain and use an estimate of the probability that an individual produces a cue during the survey. The latter is unattractive as it requires information about cue rates, and the method presented here is constructed specifically to avoid this (cf. the methodology described in Chapter 3).

Recall that $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ (Equation (4.7)) is the probability that the first cue that is detected from an individual at $\boldsymbol{x}$ is also included. Its derivation assumes that all individuals in $\mathcal{A}$ are eventually detected; that is, an individual is either detected and included or detected and not included. Thus, if $c_l \to -\infty$ (i.e., all emitted cues are detected), a survey long enough to guarantee that all individuals produce at least one cue over its duration would be required. The method presented here does not assume that all cues are detected. A survey must be long enough to guarantee that all individuals are detected (by virtue of producing at least one cue that is received with a signal strength exceeding c_l). In theory, for this to occur almost surely, the survey must be of infinite length. This allows every individual to produce as many cues as required for this to occur (thus Equation (4.5) comprises an infinite sum).

Of course, this assumption may lead to the requirement of an infeasibly long survey if there exists some $\boldsymbol{x} \in \mathcal{A}$ such that $p_c(\boldsymbol{x};\boldsymbol{\theta},c_l) \approx 0$; individuals in locations far from the detector array may need to produce a large number of cues before being detected by any detector. However, the model is robust to violations caused by having a shorter survey, as long as $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \approx 0$ for these locations. In this case, when individuals located far from the array eventually have a cue received at a signal strength exceeding c_l, there is a near-zero probability of the received signal strength also exceeding c_u. Thus, waiting for these individuals to produce a cue that is eventually detected is pointless as they are extremely unlikely to contribute a capture history that is included for analysis anyway. The result shown in the appendix immediately following this chapter is useful here: this shows that

$p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \to 0$ as distance from the array increases, providing justification for the robustness to the assumption violation caused by the survey being of finite length. This result is also seen empirically in all plots shown in Figure 4.2.

Moreover, recall from Section 4.6.1 that $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) \to 0$ as $c_u - c_l \to \infty$. The larger the difference between the two thresholds, the faster the convergence of $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ to 0, as the smaller $p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$ will be. Thus, this methodology is more robust to violations from the assumption of an infinite survey length when $c_u - c_l$ is large. However, setting c_u too high will decrease the number of capture histories eligible for analysis (as this makes it more difficult to include individuals), therefore decreasing estimator precision. It is therefore of interest to set c_l as low as feasibly possible to increase $c_u - c_l$ (as long as all cues with received signal strengths exceeding this level are detected with certainty); a further simulation study (not shown here) suggests this also leads to estimates that are more precise.

4.6.3 Model robustness of Dawson & Efford (2009)

Along with Figure 4.2(iii), the equivalence of parameter estimates obtained in Section 4.5 shows that the analysis carried out by Dawson and Efford (2009) is appropriate if it is possible to assume that all cues received above 38.4 dB (the value of the lower threshold) were detected. However, as discussed in Section 4.5, this is almost certainly not the case as background ambient noise reached levels up to 60 dB. It is likely that any signals received at strengths lower than the background noise level at the time remained undetected.

Figures 4.2(i) and (ii) show that inclusion probabilities are somewhat affected when $c_l = 60$, as in this case there are non-negligible probabilities associated with the detection of subsequent cues at some points in $\mathcal{A}$. This, however, represents a worst-case scenario; setting the lower threshold of 60 dB in the production of these plots results in the calculation of probabilities that assume *no* cues with received signals less than 60 dB were detected at any microphones.

In practice, the true scenario lies somewhere in between these two extremes: some signals between 38.4 and 60 dB were detected, and some where not. Even in the worst-case scenario, however, individuals inclusion probabilities, $p(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c})$, were only moderately affected by the inclusion of subsequent cues (Figures 4.2(i) and (ii)). It is therefore likely that very few subsequent cues were included for analysis, and so it is reasonable to believe that the estimates obtained by Dawson and Efford (2009) were largely unaffected by their failure to explicitly model their inclusion.

To formally test this it is necessary to carry out a second simulation study. It would be appropriate to simulate individuals and cues as per the steps described in Section 4.3.3 with

one modification: rather than a constant c_l being used as the threshold of detection, for each cue a background noise level could be generated from some distribution. If a received signal strength exceeds this value then the cue is detected. To match the distribution of recorded background noise levels observed by Dawson and Efford (2009; with positive skew, a median of 45.3 dB, and a 95th percentile of 52.5 dB), these could be simulated from a log-normal($\mu = 3.81, \sigma^2 = 8.04 \times 10^{-3}$) distribution. For analysis, c_l must be set at a level higher than any plausible background noise level. This means that many first cues that are picked up by the microphones must be declared undetected and thus ignored, given the definition of a 'detection' used here (i.e., a received signal strength exceeding c_l).

4.6.4 Modelling background noise

Given the framework presented in this chapter, the choice of c_l may be difficult. Recall that it is assumed all individuals eventually produce a cue that is detected over the course of the survey (and that this occurs almost surely if the survey is of infinite length). Setting c_l too high therefore requires a longer survey and potentially decreases estimator precision (as many cues must be ignored), while setting it below levels at which cues can be heard and identified violates an assumption.

If it can be assumed that cues are detected if they are received at strengths in excess of the background noise at the time, then it may be possible to fit a model with a fluctuating lower threshold at each detector, subject to their received background noise levels. This extension would allow for better modelling of the detection process; rather than simply fixing c_l at a value that is too high for all but the noisiest periods of the survey (resulting in many detections being discarded), one may allow it to decrease while detectors are not subject to high levels of background noise. The first-call models presented here therefore provide a framework under which fluctuating detectability due to background noise in acoustic surveys can be accounted for—something all existing SECR methods are unable to deal with.

4.6.5 Concluding remarks

Deriving a likelihood that incorporates all detected cues (other than only the first) from all individuals, while also allowing the use of cue-level auxiliary data, would be a useful development (see Section 7.1). Including a larger number of cues in the analysis is likely to reduce the variance of the density estimator, although obtaining estimates is likely to be substantially more computationally demanding.

First-cue models will remain useful even following the development of such a method. Depending on the species being monitored, acoustic surveys can collect large volumes of

data; attempting to estimate animal density by incorporating all detected calls (not only the first-detected calls) into a complex model may not be feasible. In such a situation it would be necessary to make a decision between (i) truncating the survey length so that the number of detected cues is manageable, or (ii) not truncating the survey and fitting a first-cue model instead, so that the number of analysed cues is no greater than the number of unique detected individuals. While the former is attractive due to all cues from all detected animals within the truncated time frame being analysed, the latter has the advantage of making use of data from a larger number of unique individuals. The compromise between analysing more information from fewer individuals or less information from more individuals, and the effect this has on estimator precision, would require further investigation.

Appendix: Limiting behaviour of subsequent-cue inclusion probabilities

Here it is shown that the probability of inclusion via a subsequent cue approaches 0 as the distance from the detector centroid increases, if and only if the upper threshold is strictly larger than the lower threshold. That is,

$$\lim_{d_\cdot(\boldsymbol{x})\to\infty} p_s(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{c}) = 0 \quad \text{if and only if } c_u > c_l. \tag{4.11}$$

For simplicity, it is assumed that all detectors are located at the same point; this is chiefly because the relationship between $d_j(\boldsymbol{x})$ and $d_\cdot(\boldsymbol{x})$ otherwise depends on the relative position of the jth detector to the detector centroid, making the above limit difficult to formally define. A justification for this being suitable is that, when considering locations far from the detector array (i.e., when $d_\cdot(\boldsymbol{x})$ is large), the spatial range of the detectors is negligible in comparison.

Approximation of $\Pr(X \geq 1)$ for binomial X

First, consider a random variable X such that

$$X \sim \text{Binomial}(n, p).$$

Here it is shown that

$$\Pr(X \geq 1) \approx np \quad \text{if } p \text{ is small.} \tag{4.12}$$

This can be achieved by considering $\Pr(X = 0)$, as follows:

$$\begin{aligned}
\Pr(X = 0) &= (1-p)^n \\
&= \sum_{i=0}^{n} \binom{n}{i} (-p)^n \\
&= 1 - np + \binom{n}{2} p^2 - \binom{n}{3} p^3 + \cdots \\
&= 1 - np + O(p^2) \quad \text{as} \quad p \to 0.
\end{aligned}$$

Thus $\Pr(X = 0)$ can be approximated by $1 - np$ when p is small, as the term $O(p^2)$ becomes negligible in comparison to np as $p \to 0$; that is,

$$\lim_{p \to 0} O(p^2)/np = 0.$$

Equation (4.12) then follows directly from the above.

Subsequent-cue inclusion probabilities

Now, for a cue emitted at location $\boldsymbol{x}$, let (i) b be the standardised difference between the expected received signal strength and the lower threshold, and (ii) a be the difference between the standardised differences between each of the upper threshold and the lower threshold, and the expected received signal strength. Thus,

$$\begin{aligned} b &= \frac{c_l - (\beta_0 - \beta_1\, d.(\boldsymbol{x}))}{\sigma_s}, \\ a &= \frac{c_u - (\beta_0 - \beta_1\, d.(\boldsymbol{x}))}{\sigma_s} - \frac{c_l - (\beta_0 - \beta_1\, d.(\boldsymbol{x}))}{\sigma_s}, \end{aligned}$$

and so,

$$a + b = \frac{c_u - (\beta_0 - \beta_1\, d.(\boldsymbol{x}))}{\sigma_s}.$$

If $c_u > c_l$, then $a > 0$, and if $c_u = c_l$ then $a = 0$. By definition $c_u \not< c_l$. These definitions hold for all detectors if they are associated with the same point in space; furthermore, they allow the probabilities of a cue's detection at strengths exceeding c_l and c_u at a single detector to be written as

$$g(d.(\boldsymbol{x}); \boldsymbol{\theta}, c_l) = 1 - \Phi(b)$$

and

$$g(d.(\boldsymbol{x}); \boldsymbol{\theta}, c_u) = 1 - \Phi(a + b),$$

respectively.

Due to the signal strength detection function having a horizontal asymptote at 0, $g(d.(\boldsymbol{x}); \boldsymbol{\theta}, c_l) \to 0$ as $d.(\boldsymbol{x}) \to \infty$. Furthermore, the probability of at least one received signal strength exceeding some threshold c is therefore the probability of a binomial ran-

dom variable being realised as greater or equal to one. Therefore—from Equation (4.12)—if $d.(\boldsymbol{x})$ is large (and therefore $g(d.(\boldsymbol{x}); \boldsymbol{\theta}, c_l)$ is small) then

$$p_c(\boldsymbol{x}; \boldsymbol{\theta}, c_l) \approx m[g(d.(\boldsymbol{x}); \boldsymbol{\theta}, c_l)] = m[1 - \Phi(b)] \tag{4.13}$$

and

$$p_c(\boldsymbol{x}; \boldsymbol{\theta}, c_u) \approx m[g(d.(\boldsymbol{x}); \boldsymbol{\theta}, c_u)] = m[1 - \Phi(a + b)]. \tag{4.14}$$

As $d.(\boldsymbol{x}) \to \infty$, the standardised difference $b \to \infty$. If all detectors are considered to be at the same location then detection probabilities across all detectors are equivalent. Therefore—along with substitution of Equations (4.13) and (4.14) for $p_c(\boldsymbol{x}; \boldsymbol{\theta}, c_l)$ and $p_c(\boldsymbol{x}; \boldsymbol{\theta}, c_u)$, respectively—this allows

$$\lim_{b \to \infty} \frac{m[1 - \Phi(a + b)]\,(1 - m[1 - \Phi(b)])}{m[1 - \Phi(b)]} = 0 \quad \text{if and only if } a > 0 \tag{4.15}$$

to be considered in place of the Equation (4.11) above without loss of generality.

This limit can be rewritten as

$$\lim_{b \to \infty} \frac{m[1 - \Phi(a + b)]\,(1 - m[1 - \Phi(b)])}{m[1 - \Phi(b)]} = \lim_{b \to \infty} \frac{[1 - \Phi(a + b)]\,(1 - m + m\Phi(b)])}{1 - \Phi(b)},$$

and, by L'Hôpital's rule,

$$= \lim_{b \to \infty} \frac{-\phi(a + b)[1 - m + m\Phi(b)] + m[1 - \Phi(a + b)]\phi(b)}{-\phi(b)}.$$

Further rearrangement gives

$$\begin{aligned}
&= \lim_{b \to \infty} \frac{\phi(a + b)[1 - m + m\Phi(b)]}{\phi(b)} - \frac{m[1 - \Phi(a + b)]\phi(b)}{\phi(b)} \\
&= \lim_{b \to \infty} \frac{\phi(a + b)[1 - m + m\Phi(b)]}{\phi(b)} - m[1 - \Phi(a + b)] \\
&= \lim_{b \to \infty} \frac{\phi(a + b)[1 - m + m\Phi(b)]}{\phi(b)}.
\end{aligned}$$

Substitution of the standard normal PDF and CDF gives

$$= \lim_{b \to \infty} \frac{(1/\sqrt{2\pi}) \exp(-(a + b)^2/2)\{1 - m + m[1 + \operatorname{erf}(b/\sqrt{2})]/2\}}{(1/\sqrt{2\pi}) \exp(-b^2/2)}$$

$$
\begin{aligned}
&= \lim_{b\to\infty} \frac{\exp(-(a+b)^2/2)}{\exp(-b^2/2)} \times \{1 - m + m[1 + \operatorname{erf}(b/\sqrt{2})]/2\} \\
&= \lim_{b\to\infty} \frac{\exp(-(a+b)^2/2)}{\exp(-b^2/2)} \times (1 - m + m), \quad \text{as } \lim_{b\to\infty} \operatorname{erf}(b/\sqrt{2}) = 1 \\
&= \lim_{b\to\infty} \frac{\exp(-(a+b)^2/2)}{\exp(-b^2/2)} \\
&= \lim_{b\to\infty} \exp\left(\frac{-(a^2 + 2ab + b^2) + b^2}{2}\right) \\
&= \lim_{b\to\infty} \exp\left(\frac{a^2}{2} - ab\right) \\
&= \begin{cases} (e^0)^b & a = 0 \\ e^{-\infty} & a > 0 \end{cases} \\
&= \begin{cases} 1 & a = 0 \\ 0 & a > 0 \end{cases} \\
&= 0, \quad \text{if and only if } a > 0.
\end{aligned}
$$

Both cases above can be supported empirically by the `mask.test()` function; in Figure 4.2 $c_u > c_l$ in all cases, and in each $p(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) \to 0$ as distance from the detector array increases. When the `lower.cutoff` argument is set to the same value as `cutoff` it is observed that $p(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{c}) \to 1$. This makes intuitive sense—given a survey of infinite length, all individuals will eventually be both detected and included (as detection without inclusion is no longer possible when $c_u = c_l$).

Chapter 5

SECR methods for cue directionality and strength heterogeneity

5.1 Introduction

Conditional on animal location, independence of detections across detectors has been a fundamental assumption made by all methods presented in this thesis thus far (and, indeed, throughout almost all existing SECR methodology). This has allowed the PMF for the ith capture history conditional on location to be specified as

$$f(\boldsymbol{\omega}_i|\boldsymbol{x}_i;\boldsymbol{\gamma}) = \frac{\prod_{j=1}^{m} f(\omega_{ij}|\boldsymbol{x}_i;\boldsymbol{\gamma})}{p(\boldsymbol{x}_i;\boldsymbol{\gamma})}, \tag{5.1}$$

where

$$f(\omega_{ij}|\boldsymbol{x}_i;\boldsymbol{\gamma}) = \begin{cases} g(d_j(\boldsymbol{x}_i);\boldsymbol{\gamma}) & \omega_{ij} = 1; \\ 1 - g(d_j(\boldsymbol{x}_i);\boldsymbol{\gamma}) & \omega_{ij} = 0. \end{cases}$$

This is equivalent to Equation (1.8). Recall the denominator in Equation (5.1) is due to capture histories of $\mathbf{0}_m$ being unobservable, the probability of which is given by $1 - p(\boldsymbol{x}_i;\boldsymbol{\gamma})$.

In many cases, this independence assumption may not be reasonable: given an individual's location, its detection at one detector may either increase or decrease the chances of detections at other detectors. Two potential mechanisms that include such dependence are as follows:

1. Individual heterogeneity: Some members of the population may be more detectable than others. For example, detection of an individual by a detector far from its associated location may indicate that this individual is particularly detectable—this suggests other detectors are more likely than usual to make a detection.

2. Spatial dependence: If one detector makes a detection, then other nearby detectors may also be likely to make a detection. For example, if two camera traps are in very close proximity, then a detection at one may suggest that a detection of the same animal at the other is likely.

In such cases, it is necessary to (i) evaluate the robustness of SECR models that assume independence, and (ii) model this dependence if estimators appear to be affected.

Finite mixture models have been proposed and used for the modelling of such individual heterogeneity in SECR (Borchers & Efford, 2008; Efford & Mowat, 2014; Obbard, Howe, & Kyle, 2010). These methods assume that there are discrete groups of individuals within a population (e.g., males and females) of which membership is unobserved. They estimate between-group heterogeneity (e.g., between males and females) while assuming that within-group homogeneity persists. However, no effort has yet been made to account for heterogeneity induced by a continuous—rather than discrete—latent variable. Likewise, spatial dependence in capture histories—beyond that incorporated by the modelling of latent animal locations—is not incorporated by any existing SECR methods.

In this chapter methods are developed to account for two scenarios (one in each of the above categories) that occur in acoustic SECR surveys that result in dependence between capture probabilities, despite conditioning on animal locations. These are (i) cue strength heterogeneity (CSH), and (ii) cue directionality (CD).

First, all methods to date have assumed that the source strength β_0 of all cues is constant, and that no between-cue heterogeneity exists. While this is a reasonable assumption for the *A. lightfooti* data, these models are unlikely to appropriately generalise to many other species. It may be more realistic to consider each source strength a realisation of some random variable with an associated distribution characterised by parameters that require estimation. This leads to heterogeneity in detectability, as individuals producing louder cues are more likely to be detected.

Second, almost all SECR methods assume that the probability of detection at a detector is solely a function of its Euclidean distance from the individual in question. When detection is via direct observation of the individual (e.g., due to physical capture or using camera traps) then this assumes that individual home ranges are circular; when detection is via cues produced by individuals propagating through space then this assumes that the cues are

omnidirectional. Both are unlikely to hold in many scenarios. SECR methods have recently been developed to allow for the use of an 'ecological' (rather than Euclidean) distance metric based on landscape connectivity (Sutherland, Fuller, & Royle, 2014), thus relaxing this assumption. This requires knowledge of the habitat so that estimation of the ease with which individuals are able to travel through components of its structure can be estimated; travelling large Euclidean distances through habitat types that are conducive to animal movement may nevertheless incur a small ecological cost. If individuals are detected from their cues, then a similar approach could be taken in order to estimate the ease with which these signals propagate throughout the habitat. Consider acoustic detectors positioned on the edge of a dense forest leading to open plains: acoustic signals are likely more difficult if they are produced at a location within the former.

On the other hand, however, it may be the nature of the cues themselves—rather than that of the local environment—that results in a departure from omnidirectional propagation. Vocalisations of many species are produced in such a way that they are more easily detectable in a particular direction. This is often the case for animals that rely on echolocation for navigation (e.g., delphinids and microbats). Patricelli, Dantzker, and Bradbury (2007) suggest that vocalisation directionality in some species may be related to the intended function of the communication: those that are intended to be heard by many individuals (e.g., advertisement and alarm calls) are omnidirectional, while those intended for a specific individual (e.g., related to courtship) are directional, thus maximising the chances that it is heard by the intended recipient while minimising eavesdropping.

In the context of acoustic SECR surveys, this means that the probability of detecting an animal's vocalisation may not depend only on the distance it is from a detector, but also the direction in which it is emitted. A detector a great distance away may detect a cue that is emitted in its direction, while, on the other hand, a nearby detector may not detect a cue that is emitted in the opposite direction. Thus, conditional on the location of an individual, a detection made by one microphone may suggest that other nearby detectors are also likely to make a detection, as there is a good chance that the individual is facing the right direction for this to occur. Likewise, it may suggest that detectors on the other side of the individual are less likely to make a detection.

An acoustic detector array is not capable of directly observing the directions in which detected cues were emitted. Directionality and distance are confounded when considering the detection of a single cue at a single microphone—a detected cue might have been produced at a nearby location in the opposite direction to the detector, or at a location far from the detector but in the direction of the detector. However, it may be possible

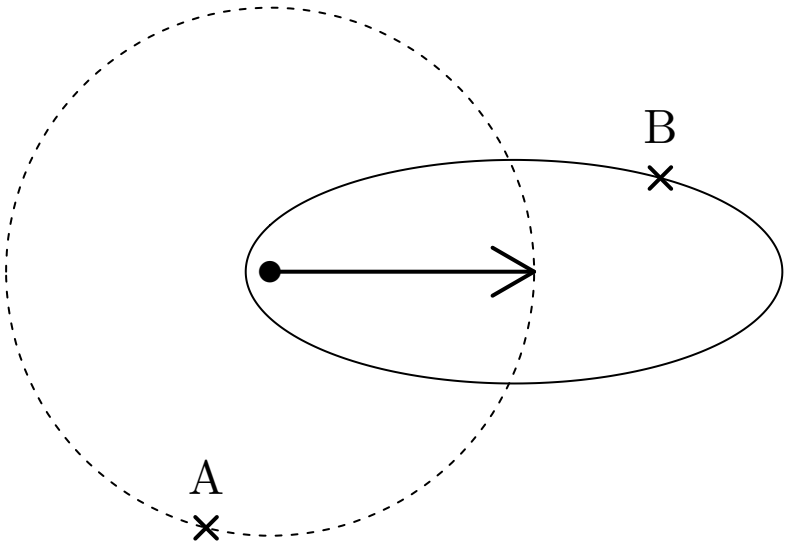


Figure 5.1 *A depiction of a directional call. The solid point is the location of the source of a cue, emitted in the direction shown by the arrow. Crosses show the locations of two detectors. All locations on the dashed, circular contour have the same expected TOA, as this is not affected by cue directionality. All locations on the solid contour have the same expected received signal strength; the cue is most easily detected by detectors in line with its direction. Suppose that Detector A received the signal earlier than Detector B, but at a weaker strength. The TOA information indicates that it was emitted at a location closer to Detector B; however, despite this, it was detected at a higher signal strength by Detector A. It is then logical to conclude that the cue was likely to have been emitted in the approximate direction of Detector A, and thus information is available about cue directionality from the signal strength and TOA auxiliary data together.*

to distinguish between these two cases when signal strength and TOA data are considered alongside the capture histories because signal strengths are affected by the cue's direction, but TOAs are not (Figure 5.1).

In this chapter SECR methodology is presented that accounts for both sources of detection dependence discussed above. In Section 5.2 methods are described that are able to estimate CD (Section 5.2.1) and CSH (Section 5.2.2) in acoustic SECR surveys. Computational details relevant to each are also presented, while examples of the fitting of these models using `admbsecr` are shown in Section 5.3. Simulation studies are conducted showing that the derived estimators are appropriate, and that ignoring these effects can result in severely biased estimation (Section 5.4).

5.2 Methodology

This section separately presents methods for the incorporation of CD (Section 5.2.1) and CSH (Section 5.2.2). Although it is possible for both to be integrated into the same model, this is not considered here. In both cases, dependence between the detections of a cue across the detectors can be thought of as being induced by an unobserved latent variable—the direction it was emitted (for the former) or its source strength (for the latter).

The capture histories and auxiliary data considered here have cues as the unit of detection and independence is assumed between all detected cues. Thus, these models are not appropriate for data in which there is dependence between cues (from the same individual, for example). Nevertheless, the extensions described here have the potential to be incorporated into existing methods (such as those described in Chapters 3 and 4) and others yet to be developed to give models that can appropriately estimate the parameters of interest from real acoustic detection data. See Section 5.5.3 for further discussion and Section 7.1 for a potential example.

Sections 5.2.1 and 5.2.2 involve the generalisation of the likelihood presented in Chapter 2 for the incorporation of auxiliary data (Equation (2.2)).

5.2.1 Cue directionality

In this section $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are assumed to contain signal strength and TOA information, respectively. While the methods outlined below allow estimation without TOA information (and an extension to situations without signal strength data is straightforward; see Section 5.5.5) it is incorporated here as these data appear particularly informative about cue directionality (Figure 5.1). The addition of any other auxiliary data types that are not affected by cue directionality can be achieved using the methods described in Chapter 2.

Let u_i be the bearing (in radians) of the direction in which the ith detected cue was emitted, let u denote a generic direction of a cue, and recall that $b_j(\boldsymbol{x}_i)$ gives the bearing from the ith detected cue to the jth detector. Thus $u_i - b_j(\boldsymbol{x}_i)$ provides the *bearing discrepancy* between the ith individual and the jth trap. For example, this is equal to 0 if the jth trap is directly in line with the direction of the emitted cue, and equal to π if it is emitted in the opposite direction.

A detection is therefore most likely if the bearing discrepancy is near 0, and least likely if it is near π. Recall from Section 2.2.3, Chapter 2, that the expected received signal strength is assumed to depend on three parameters, these being β_0, β_1, and σ_s. Specifically, β_1 controls the rate at which the expected received signal strength decreases as the distance

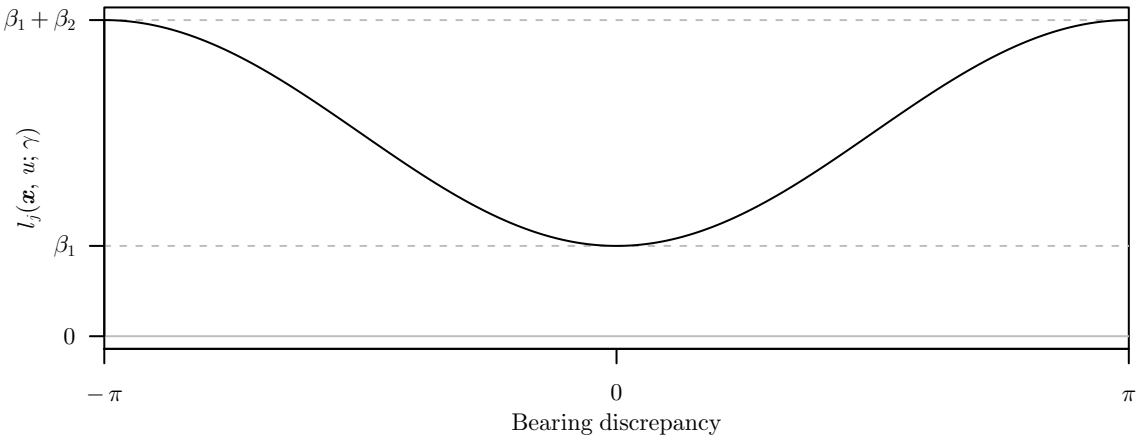


Figure 5.2 *The assumed relationship between bearing discrepancy and signal strength loss rate, $l_j(\boldsymbol{x}_i, u_i; \boldsymbol{\gamma})$. Signal strengths decay at the slowest rate in the direction the cue was emitted (a bearing discrepancy of 0), and fastest in the opposite direction to this (a bearing discrepancy of π or $-\pi$). The parameter β_2 measures the difference between these two rates.*

between the cue's source location and the detector increases. When cues are directional it is sensible to consider that this varies depending on the bearing discrepancy; that is, the received signal strength should be assumed to decrease slowest in the direction that the call was emitted. The exact functional form of this relationship assumed here is

$$l_j(\boldsymbol{x}_i, u_i; \boldsymbol{\gamma}) = \beta_1 - \frac{\beta_2[\cos(u_i - b_j(\boldsymbol{x}_i)) - 1]}{2}, \tag{5.2}$$

where $l_j(\cdot)$ provides the *signal strength loss rate* in the direction of the jth microphone, and β_1 is now defined as the signal strength loss rate in the direction that the cue was emitted (Figure 5.2). The parameter β_2 provides the difference in this rate between bearing discrepancies of 0 and π (i.e., the emission direction and the direction opposite to this, respectively), and so the special case $\beta_2 = 0$ collapses to the nondirectional model, where $l_j(\boldsymbol{x}_i, u_i; \boldsymbol{\gamma}) = \beta_1$ for all j. Conditional on cue direction, the particular relationship between the expected received signal strength and the distance to the detector is a modelling decision; the identity-link, log-link, and spherical-spreading options presented in Section 2.2.3, Chapter 2 are three possible choices, with $l_j(\boldsymbol{x}_i, u_i; \boldsymbol{\gamma})$ replacing β_1 in both Equations (2.11) and (2.13).

However, despite affecting expected received signal strengths (and therefore detection probabilities), recall that the direction in which each cue is emitted remains unobserved. These can instead be considered realisations of latent variables. One plausible assumption (used throughout this chapter) is that an individual is equally likely to emit its cue in any

direction, and this provides the PDF

$$f(u) = \frac{1}{2\pi}, \quad u \in [0, 2\pi). \tag{5.3}$$

Alternatively—although not considered further here—it may be hypothesised that members of the population prefer to face a particular direction; in such a case parameters characterising some circular distribution (e.g., a von Mises or wrapped Cauchy distribution) can be estimated in order to infer this favoured direction and the population variance around it.

With the addition of these latent variables many functions defined previously in this thesis are no longer available in closed form, and must instead be evaluated via marginalisation over all possible directions in which cues may have been emitted. Let the subscript u indicate that a function is conditional on the direction of a cue. For example, using the signal strength detection function (Equation (2.11)), the probability of a cue emitted at $\boldsymbol{x}$ being detected by the jth detector, conditional call direction, is given by

$$g_u(d_j(\boldsymbol{x})|u;\boldsymbol{\gamma}) = 1 - \Phi\left(\frac{c - h^{-1}(\beta_0 - l_j(\boldsymbol{x}, u;\boldsymbol{\gamma})d_j(\boldsymbol{x}))}{\sigma_s}\right),$$

and the marginal probability (unconditional on cue directionality) is therefore

$$g(d_j(\boldsymbol{x});\boldsymbol{\gamma}) = \int_0^{2\pi} g_u(d_j(\boldsymbol{x})|u;\boldsymbol{\gamma})\, f(u)\, \mathrm{d}u.$$

Likewise, the probabilities of detection by any detector in the array can be specified in the conditional and unconditional forms

$$p_u(\boldsymbol{x}|u;\boldsymbol{\gamma}) = 1 - \prod_{j=1}^{m}[1 - g_u(d_j(\boldsymbol{x})|u;\boldsymbol{\gamma})]$$

and

$$p(\boldsymbol{x};\boldsymbol{\gamma}) = \int_0^{2\pi} p_u(\boldsymbol{x}|u;\boldsymbol{\gamma})\, f(u)\, \mathrm{d}u, \tag{5.4}$$

respectively.

This also allows specification of the PMF of the ith capture history, now conditional on both the location at which it was emitted along with its direction:

$$f(\boldsymbol{\omega}_i|\boldsymbol{x}_i, u_i;\boldsymbol{\gamma}) = \frac{\prod_{j=1}^{m} f(\omega_{ij}|\boldsymbol{x}_i, u_i;\boldsymbol{\gamma})}{p_u(\boldsymbol{x}_i|u_i;\boldsymbol{\gamma})}, \tag{5.5}$$

where

$$f(\omega_{ij}|\boldsymbol{x}_i, u_i; \boldsymbol{\gamma}) = \begin{cases} g_u(d_j(\boldsymbol{x}_i)|u_i; \boldsymbol{\gamma}) & \omega_{ij} = 1; \\ 1 - g_u(d_j(\boldsymbol{x}_i)|u_i; \boldsymbol{\gamma}) & \omega_{ij} = 0. \end{cases}$$

The joint PDF of the received signal strengths corresponding to the ith detected cue—conditional on its capture history, location, and direction—involves simple extensions to Equations (2.6) and (2.12); that is,

$$f(\boldsymbol{y}_{i\cdot 1}|\boldsymbol{\omega}_i, \boldsymbol{x}_i, u_i; \boldsymbol{\gamma}) = \prod_{\{j:\omega_{ij}=1\}} f(y_{ij1}|\boldsymbol{x}_i, u_i; \boldsymbol{\gamma}) \tag{5.6}$$

and

$$f(y_{ij1}|\boldsymbol{x}_i, u_i; \boldsymbol{\gamma}, c) = \frac{\phi([y_{ijk} - h^{-1}(\beta_0 - l_j(\boldsymbol{x}_i, u_i; \boldsymbol{\gamma})\, d_j(\boldsymbol{x}_i))]/\sigma_s)}{\sigma_s\{1 - \Phi([c - h^{-1}(\beta_0 - l_j(\boldsymbol{x}_i, u_i; \boldsymbol{\gamma})\, d_j(\boldsymbol{x}_i))]/\sigma_s)\}}.$$

Finally, it is necessary to now generalise the likelihood itself. From Equations (2.2) and (2.3)—specifically incorporating $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ as signal strength and TOA auxiliary information, respectively—this is

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{\Omega}, \boldsymbol{Y}) &= f(n; D, \boldsymbol{\gamma})\, f(\boldsymbol{\Omega}, \boldsymbol{Y}|n; \boldsymbol{\gamma}, \boldsymbol{\psi}) \\ &= f(n; D, \boldsymbol{\gamma}) \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{y}_{i\cdot 1}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{y}_{i\cdot 2}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}_i. \end{aligned}$$

However, as per the above, $\boldsymbol{Y}_1$, and $\boldsymbol{\Omega}$ depend on the direction of the ith cue. Thus, u_i must be marginalised over within this integrand:

$$= f(n; D, \boldsymbol{\gamma}) \prod_{i=1}^{n} \int_{\mathcal{A}} \int_{0}^{2\pi} f(\boldsymbol{y}_{i\cdot 1}|\boldsymbol{\omega}_i, \boldsymbol{x}_i, u_i; \boldsymbol{\psi})\, f(\boldsymbol{y}_{i\cdot 2}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i, u_i; \boldsymbol{\gamma})\, f(u_i)\, f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}u_i\, \mathrm{d}\boldsymbol{x}_i \tag{5.7}$$

$$= f(n; D, \boldsymbol{\gamma}) \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{y}_{i\cdot 2}|\boldsymbol{\omega}_i, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma}) \int_{0}^{2\pi} f(\boldsymbol{y}_{i\cdot 1}|\boldsymbol{\omega}_i, \boldsymbol{x}_i, u_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i, u_i; \boldsymbol{\gamma})\, f(u_i)\, \mathrm{d}u_i\, \mathrm{d}\boldsymbol{x}_i. \tag{5.8}$$

Neither the conditional PDF for $\boldsymbol{y}_{i\cdot 2}$ nor the marginal PDF for $\boldsymbol{x}_i$ depend on u_i, allowing for their above removal from the inner integrand.

The three remaining terms in this inner integrand have been specified in Equations (5.6), (5.5), and (5.3), respectively. With TOAs unaffected by cue directionality, the first term in the outer integrand remains as per Equation (2.14). The marginal PDF for the location of the ith cue remains as per Equation (1.4), but using the DPS due to cue directionality, as specified above in Equation (5.4). That is,

$$\begin{aligned} f(\boldsymbol{x}_i;\boldsymbol{\gamma}) &= \frac{p(\boldsymbol{x}_i;\boldsymbol{\gamma})}{\int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{x}} \\ &= \frac{\int_0^{2\pi} p_u(\boldsymbol{x}_i|u_i;\boldsymbol{\gamma})\,f(u_i)\,\mathrm{d}u_i}{\int_{\mathcal{A}}\int_0^{2\pi} p_u(\boldsymbol{x}|u;\boldsymbol{\gamma})\,f(u)\,\mathrm{d}u\,\mathrm{d}\boldsymbol{x}}, \end{aligned}$$

where the denominator above specifies the ESA.

The likelihood for a model incorporating cue directionality has now been fully specified, its maximisation providing $\widehat{\boldsymbol{\theta}}$.

Approximating the marginalisation of cue directions can be achieved numerically via quadrature. That is, n_q evenly spaced points are selected over $[0, 2\pi)$, and let $\mathcal{D}$ represent this set. The inner integral of the likelihood can then be approximated using

$$\begin{aligned} \int_0^{2\pi} f(\boldsymbol{y}_{i\cdot 1}|\boldsymbol{\omega}_i,\boldsymbol{x}_i,u_i;\boldsymbol{\psi})\,f(\boldsymbol{\omega}_i|\boldsymbol{x}_i,u_i;\boldsymbol{\gamma})\,f(u_i)\,\mathrm{d}u_i \approx \frac{2\pi}{n_q}\sum_{u_i\in\mathcal{D}} f(\boldsymbol{y}_{i\cdot 1}|\boldsymbol{\omega}_i,\boldsymbol{x}_i,u_i;\boldsymbol{\psi}) \\ f(\boldsymbol{\omega}_i|\boldsymbol{x}_i,u_i;\boldsymbol{\gamma})\,f(u_i). \end{aligned} \tag{5.9}$$

The same approach can be used for the calculation of $p(\boldsymbol{x};\boldsymbol{\gamma})$ (Equation (5.4)).

5.2.2 Cue strength heterogeneity

In this section, let $\boldsymbol{Y}$ solely represent signal strength data for simplicity (although including any other kinds of additional auxiliary data is straightforward), and so subscripts referring to the type of auxiliary information that $\boldsymbol{Y}$ represents are omitted. Rather than assuming that the source strength of all cues is constant at β_0, let the ith detected cue have a source strength of v_i. Here it is assumed that

$$v_i \sim \mathrm{N}(\beta_0, \sigma_v), \tag{5.10}$$

although estimation remains possible with other distributional assumptions. Here β_0 is redefined as the *mean* population source strength, and so the methods used to model received signal strengths in previous chapters fall under the special case $\sigma_v = 0$. Let σ_v be

an element of the parameter vector $\boldsymbol{\gamma}$ for models incorporating CSH.

A similar approach to what was used for cue directionality (Section 5.2.1) could be employed here, whereby the likelihood is given by marginalisation over v_i:

$$L(\boldsymbol{\theta}; \boldsymbol{\Omega}, \boldsymbol{Y}) = f(n; D, \boldsymbol{\gamma}) \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{x}_i \tag{5.11}$$

$$= f(n; D, \boldsymbol{\gamma}) \prod_{i=1}^{n} \int_{\mathcal{A}} f(\boldsymbol{x}_i; \boldsymbol{\gamma}) \int_{-\infty}^{+\infty} f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i, v_i; \boldsymbol{\gamma})\, f(v_i; \boldsymbol{\gamma})\, \mathrm{d}v_i\, \mathrm{d}\boldsymbol{x}_i, \tag{5.12}$$

and with the inner integral again approximated numerically by quadrature. However—under the normality assumption above—it was thought that it may have been possible to specify $f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i; \boldsymbol{\gamma})$ (in Equation (5.11)) in closed form. Without resorting to numerical approximation for its evaluation, the calculation of the likelihood function would improve in both accuracy and its computational efficiency. Deriving a general closed-form expression for the inner integral eventually proved unsuccessful, but the steps taken to attempt this are described below. These result in a different (though equivalent) specification of the inner integral that is more accurate and more efficient to evaluate than approximation of the inner integral in Equation (5.12) in some cases.

Alternative derivation

Let e_{ij} be the error associated with the measurement of the received signal strength of the ith cue detected cue at the jth detector. As per Section 2.2.3 (and specifically Equation (2.10)),

$$e_{ij} \sim \mathrm{N}(0, \sigma_s). \tag{5.13}$$

These measurement errors are assumed to be independent across all m detectors. It is further assumed—for the sake of simplicity—that the expected received signal strength decreases linearly with distance (but once again these methods can be readily extended to the log-linear or spherical-spreading relationships described in Section 2.2.3), and therefore

$$y^*_{ij} = v_i - \beta_1\, d_j(\boldsymbol{x}_i) + e_{ij}. \tag{5.14}$$

Recall that the * superscript indicates that this received signal strength is not necessarily observed—it may be too low to warrant detection at the detector. However, if $y_{ijk} > c$ it is observed with certainty. Thus, from Equations (5.10), (5.13), and (5.14),

$$y^*_{ij} | \boldsymbol{x}_i \sim \mathrm{N}(\beta_0 - \beta_1\, d_j(\boldsymbol{x}_i), \sigma_v^2 + \sigma_s^2).$$

Furthermore, the received signal strengths across all m detectors have a multivariate normal distribution; that is,

$$\boldsymbol{y}_i^*|\boldsymbol{x}_i \sim \mathrm{N}_m(\beta_0 - \beta_1\, \boldsymbol{d}(\boldsymbol{x}_i), \boldsymbol{\Sigma}_i), \tag{5.15}$$

where the vector $\boldsymbol{d}(\boldsymbol{x}_i) = (d_1(\boldsymbol{x}), \cdots, d_m(\boldsymbol{x}))$ provides distances between $\boldsymbol{x}$ and all detectors, and $\boldsymbol{\Sigma}_i$ is the variance-covariance matrix of the vector of random variables $\boldsymbol{y}_i^*|\boldsymbol{x}_i$. The latter has diagonal elements

$$(\boldsymbol{\Sigma}_i)_{jj} = \mathrm{Var}(y_{ij}) = \sigma_v^2 + \sigma_s^2, \quad \text{for all } j \tag{5.16}$$

and offdiagonal elements

$$(\boldsymbol{\Sigma}_i)_{jj'} = \mathrm{Cov}(y_{ij}, y_{ij'}) = \sigma_v^2, \quad \text{for all } j \neq j'; \tag{5.17}$$

the latter due to y_{ij} and $y_{ij'}$ only sharing the common variance component σ_v^2. This can be formally derived, but this is not shown here for brevity and due to the intuitive nature of the result.

The conditional PDF of the random vector $\boldsymbol{y}_i^*|\boldsymbol{x}_i$ is therefore given by

$$f(\boldsymbol{y}_i^*|\boldsymbol{x}_i;\boldsymbol{\gamma}) = (2\pi)^{-m/2}|\boldsymbol{\Sigma}_i|^{-1/2}\exp(\{\boldsymbol{y}_i - [\beta_0 - \beta_1\boldsymbol{d}(\boldsymbol{x}_i)]\}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\{\boldsymbol{y}_i - [\beta_0 - \beta_1\boldsymbol{d}(\boldsymbol{x}_i)]\}/2), \tag{5.18}$$

the multivariate normal PDF.

Next, let $\boldsymbol{\omega}_i^+$ and $\boldsymbol{\omega}_i^-$ partition the vector $\boldsymbol{\omega}_i$, where the vector $\boldsymbol{\omega}_i^+$ contains elements corresponding to detectors that detected the ith detected cue, and the vector $\boldsymbol{\omega}_i^-$ contains elements corresponding to detectors that did not detect the ith detected cue. Likewise, let $\boldsymbol{y}_i^+$ correspond to received signal strengths at microphones that made a detection. The vector $\boldsymbol{y}_i^-$ is not observed as its elements correspond to signal strength that did not exceed c, and were therefore potentially too weak for a measurement to be recorded (recall that if $\omega_{ij} = 0$ then the corresponding auxiliary data are not necessarily observed and do not appear in the likelihood; see Equation (2.6)). The joint PDF that requires derivation can then be respecified as

$$\begin{aligned} f(\boldsymbol{\omega}_i, \boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\gamma}) &= f(\boldsymbol{\omega}_i^+, \boldsymbol{\omega}_i^-, \boldsymbol{y}_i^+|\boldsymbol{x}_i;\boldsymbol{\gamma}) \\ &= f(\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+|\boldsymbol{x}_i;\boldsymbol{\gamma})\, f(\boldsymbol{\omega}_i^-|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i;\boldsymbol{\gamma}) \end{aligned} \tag{5.19}$$

Given a detection by the jth detector, the corresponding received signal strength must be greater than c. Thus, conditional on $\omega_{ij} = 1$, y_{ij} is a realisation from a truncated normal

distribution (as per Equation (2.12)). As an extension, conditional on $\boldsymbol{\omega}_i^+ = \mathbf{1}$, the vector $\boldsymbol{y}_i^+$ is a realisation from a truncated multivariate normal distribution, with lower truncation occurring at c on every dimension. This gives the conditional PDF

$$f(\boldsymbol{y}_i^+|\boldsymbol{\omega}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma}) = \frac{f(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i; \boldsymbol{\gamma})}{\int_{\mathcal{G}_i^+} f(\boldsymbol{y}_i^{*+\prime}|\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{y}^{*+\prime}},$$

where $\boldsymbol{y}_{ij}^{*+\prime}$ is used as a dummy variable for $\boldsymbol{y}_i^{*+}$, and the domain of integration in the denominator is given by the z_i-dimensional space $\mathcal{G}_i^+ = [c, \infty)^{z_i}$. The above is therefore a z_i-dimensional integral; recall that z_i is the number of detectors that detected the ith cue. The numerator and denominator are both available from Equation (5.18). The denominator provides the probability of detecting the cue at all detectors that made a detection—that is, the probability of each having a received signal strength in the interval $[c, \infty)$. Additionally, this is precisely what is required for the calculation of $f(\boldsymbol{\omega}_i^+|\boldsymbol{x}_i; \boldsymbol{\gamma})$; therefore

$$\begin{aligned} f(\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+|\boldsymbol{x}_i; \boldsymbol{\gamma}) &= f(\boldsymbol{\omega}_i^+|\boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{y}_i^+|\boldsymbol{\omega}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma}) \\ &= \int_{\mathcal{G}_i^+} f(\boldsymbol{y}_i^{*+\prime}|\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{y}^{*+\prime} \frac{f(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i; \boldsymbol{\gamma})}{\int_{\mathcal{G}_i^+} f(\boldsymbol{y}_i^{*+\prime}|\boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{y}^{*+\prime}} \\ &= f(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i; \boldsymbol{\gamma}). \end{aligned} \tag{5.20}$$

This is available in closed form directly from the multivariate normal PDF (i.e., as provided in Equation (5.18), but only using elements of $\boldsymbol{y}_i$ that correspond to detectors that successfully made a detection).

To complete the derivation of the joint density (Equation (5.19)), all that remains is to specify $f(\boldsymbol{\omega}_i^-|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma})$. This is the probability of nondetection at the detectors that did not detect the cue, conditional on its detection and the received signal strengths at the other detectors. The dependence between $\boldsymbol{\omega}_i^-$ and $\boldsymbol{y}_i^+$ is induced via the latent source signal strength, v_i. For example, if received signal strengths were higher than one would typically expect across all detectors that made a detection, then this would indicate that the emitted source strength v_i was high—probabilities of nondetection would therefore be smaller than usual.

The nondetection of the ith cue at the jth detector is due to $y_{ij}^* < c$. Thus, the probability of nondetection at the detectors that did not detect the cue is

$$f(\boldsymbol{\omega}_i^-|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma}) = \int_{\mathcal{G}_i^-} f(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{y}^{*-}, \tag{5.21}$$

where the domain of integration above is given by the set $\mathcal{G}_i^- = (-\infty, c)^{m-z_i}$.

The integrand is the joint PDF of some elements in the multivariate normal random vector $\boldsymbol{y}_i^*$ conditional on the remaining elements, and this requires derivation. Let the elements of both the expectation vector and the variance-covariance matrix of the random vector $\boldsymbol{y}_i^*|\boldsymbol{x}_i$ be reordered so that

$$\mathrm{E}(\boldsymbol{y}_i^*|\boldsymbol{x}_i) = \begin{pmatrix} \mathrm{E}(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i) \\ \mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{x}_i) \end{pmatrix} = \beta_0 - \beta_1 \begin{pmatrix} \boldsymbol{d}(\boldsymbol{x}_i^+) \\ \boldsymbol{d}(\boldsymbol{x}_i^-) \end{pmatrix},$$

where $\boldsymbol{x}_i^+$ and $\boldsymbol{x}_i^-$ provide the locations of the detectors that did and did not detect the ith cue, respectively, and

$$\mathrm{Var}(\boldsymbol{y}_i^*|\boldsymbol{x}_i) = \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{++} & \boldsymbol{\Sigma}_i^{+-} \\ (\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}} & \boldsymbol{\Sigma}_i^{--} \end{pmatrix},$$

where (i) $\boldsymbol{\Sigma}_i^{++}$ is a $z_i \times z_i$ submatrix containing variances of received signal strengths at detectors that made a detection and covariances between these, (ii) $\boldsymbol{\Sigma}_i^{+-}$ is a $z_i \times (m - z_i)$ submatrix containing covariances of received signal strengths between each detector that did make a detection and each that did not, and (iii) $\boldsymbol{\Sigma}_i^{--}$ is a $(m-z_i) \times (m-z_i)$ submatrix containing variances of received signal strengths at detectors that did not make a detection and covariances between these. From Equations (5.16) and (5.17) both $\boldsymbol{\Sigma}_i^{++}$ and $\boldsymbol{\Sigma}_i^{--}$ are matrices with diagonal elements $\sigma_v^2 + \sigma_s^2$ and offdiagonal elements σ_v^2, while all elements of $\boldsymbol{\Sigma}_i^{+-}$ are σ_v^2.

The distribution of the conditional random variable $\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i$ is also multivariate normal. The derivation of its expectation vector and variance-covariance matrix is possible due to the results of Eaton (1983, pp. 116–119), which provide

$$\mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i) = \mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{x}_i) + (\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}}(\boldsymbol{\Sigma}_i^{++})^{-1}[\boldsymbol{y}_i^{*+} - \mathrm{E}(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i)] \tag{5.22}$$

and

$$\mathrm{Var}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i) = \boldsymbol{\Sigma}_i^{-|+} = \boldsymbol{\Sigma}_i^{--} - (\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}}(\boldsymbol{\Sigma}_i^{++})^{-1}\boldsymbol{\Sigma}_i^{+-}. \tag{5.23}$$

In particular, the inverse of $\boldsymbol{\Sigma}_i^{++}$ is required in order to calculate $\mathrm{Var}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i)$. While implementations[1] of various algorithms are available to compute matrix inverses, computation time can be reduced by recognising that—in this specific case—it will always have a particular structure. For a $d \times d$ variance-covariance matrix $\boldsymbol{A}$ with all diagonal

[1]For example, the inverse of the matrix `A` can be calculated using `solve(A)` in R or `inv(A)` in ADMB.

elements equal to a and all offdiagonal elements equal to b,

$$\boldsymbol{A}^{-1} = \frac{1}{a^2 + (d-2)ab - (d-1)b^2}\boldsymbol{B}, \tag{5.24}$$

where the $d \times d$ matrix $\boldsymbol{B}$ has all diagonal elements equal to $a + b(d-2)$ and all offdiagonal elements equal to $-b$. Thus, $(\boldsymbol{\Sigma}_i^{++})^{-1}$ is a $z_i \times z_i$ matrix whereby diagonal elements are given by

$$\begin{aligned}(\boldsymbol{\Sigma}_i^{++})^{-1}_{jj} &= \frac{\sigma_v^2 + \sigma_s^2 + (z_i - 2)\sigma_v^2}{(\sigma_v^2 + \sigma_s^2)^2 + (z_i - 2)(\sigma_v^2 + \sigma_s^2)\sigma_v^2 - (z_i - 1)\sigma_v^2} \\ &= \frac{\sigma_s^2 + (z_i - 1)\sigma_v^2}{\sigma_s^4 + z_i\sigma_v^2\sigma_s^2},\end{aligned} \tag{5.25}$$

and offdiagonal elements are given by

$$(\boldsymbol{\Sigma}_i^{++})^{-1}_{jj'} = \frac{-\sigma_v^2}{\sigma_s^4 + z_i\sigma_v^2\sigma_s^2}, \quad \text{for all } j \neq j'. \tag{5.26}$$

As an aside, this is also the variance-covariance matrix required for the calculation of the multivariate normal PDF $f(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i;\boldsymbol{\gamma})$ (using that specified in Equation (5.18), but only including elements of $\boldsymbol{y}_i$ corresponding to detectors that made a detection), and this requires calculation of both its inverse and its determinant. The inverse is given above, and the determinant of a matrix with this structure is

$$|\boldsymbol{\Sigma}_i^{++}| = \sigma_s^{2(z_i-1)}(\sigma_s^2 + z_i\sigma_v^2). \tag{5.27}$$

Thus both are directly available for efficient calculation of $f(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i;\boldsymbol{\gamma})$ without relying on their algorithmic computation.

The expectation vector (Equation (5.22)) and variance-covariance matrix (Equation (5.23)) both require calculation of $(\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}}(\boldsymbol{\Sigma}_i^{++})^{-1}$. From Equation (5.17),

$$(\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}} = \sigma_v^2\mathbf{1}_{m-z_i,z_i},$$

where $\mathbf{1}_{a,b}$ represents an $a \times b$ matrix with all elements equal to 1. Therefore,

$$(\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}}(\boldsymbol{\Sigma}_i^{++})^{-1} = \frac{\sigma_v^2}{\sigma_s^2 + z_i\sigma_v^2}\mathbf{1}_{m-z_i,m-z_i}.$$

Postmultiplication by the vector $\boldsymbol{y}_i^{*+} - \mathrm{E}(\boldsymbol{y}_i^{*+}|\boldsymbol{x}_i)$ followed by addition of $\mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{x}_i)$, as per

Equation (5.22), provides

$$\mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i) = \mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{x}_i) + \frac{\sigma_v^2 \sum_{\{j:\omega_{ij}=1\}} [y_{ij} - \mathrm{E}(y_{ij}^*|\boldsymbol{x}_i)]}{\sigma_s^2 + z_i\sigma_v^2}. \tag{5.28}$$

Likewise, for the variance-covariance matrix, postmultiplication by the matrix $\boldsymbol{\Sigma}_i^{+-}$ gives

$$\begin{aligned}(\boldsymbol{\Sigma}_i^{+-})^{\mathrm{T}}(\boldsymbol{\Sigma}_i^{++})^{-1}\boldsymbol{\Sigma}_i^{+-} &= \frac{z_i\sigma_v^4\sigma_s^2}{\sigma_s^4 + z_i\sigma_v^2\sigma_s^2}\mathbf{1}_{m-z_i,m-z_i} \\ &= \frac{z_i\sigma_v^4}{\sigma_s^2 + z_i\sigma_v^2}\mathbf{1}_{m-z_i,m-z_i}.\end{aligned}$$

Thus, from Equation (5.23), diagonal elements of $\mathrm{Var}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i)$, that is, the variance-covariance matrix, are

$$\begin{aligned}(\boldsymbol{\Sigma}_i^{-|+})_{jj'} &= \sigma_s^2 + \sigma_v^2 - \frac{z_i\sigma_v^4}{\sigma_s^2 + z_i\sigma_v^2} \\ &= \frac{\sigma_s^4 + (z_i+1)\sigma_s^2\sigma_v^2}{\sigma_s^2 + z_i\sigma_v^2},\end{aligned} \tag{5.29}$$

while offdiagonal elements are

$$\begin{aligned}(\boldsymbol{\Sigma}_i^{-|+})_{jj} &= \sigma_v^2 - \frac{z_i\sigma_v^4}{\sigma_s^2 + z_i\sigma_v^2} \\ &= \frac{\sigma_s^2\sigma_v^2}{\sigma_s^2 + z_i\sigma_v^2}, \quad \text{for all } j \neq j'.\end{aligned} \tag{5.30}$$

Equations (5.28), (5.29), and (5.30) therefore fully specify the conditional expectation vector (Equation (5.22)) and the variance-covariance matrix (Equation (5.23)) in a way that no longer requires any algorithmic matrix inversion whatsoever. This is particularly useful in situations where cues are detected by many detectors, resulting in $\boldsymbol{\Sigma}_i^{++}$ being a large matrix—inversion of which (for each detected cue) may otherwise pose a considerable computational challenge.

One additional point to note is that, from $\mathrm{Var}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i)$, all offdiagonal elements of the corresponding correlation matrix, $\mathrm{Cor}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i) = \boldsymbol{\rho}_i^{-|+}$, are given by

$$\boldsymbol{\rho}_{jj'}^{-|+} = \frac{\sigma_v^2}{\sigma_s^2 + (z_i+1)\sigma_s^2}, \quad \text{for all } j \neq j'. \tag{5.31}$$

Let this correlation simply be represented as $\rho_i^{-|+}$ below.

The conditional PDF $f(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma})$ is therefore available as the standard multivariate normal PDF, using $\mathrm{E}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i)$ (Equation (5.28)) and $\mathrm{Var}(\boldsymbol{y}_i^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i)$ (Equations (5.29) and (5.30)) as the expectation vector and variance-covariance matrix, respectively.

In order to fully specify the required joint density in Equation (5.19) this conditional PDF must be integrated over to give $f(\boldsymbol{\omega}_i^-|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma})$ (Equation (5.21)). Definite integrals over a multivariate normal PDF—such as this—are sometimes referred to as multivariate normal CDFs (e.g., Curnow & Dunnett, 1962). However, these are not available in closed form; indeed Kotz, Balakrishnan, and Johnson (2000) state that "there does not appear to be any simple procedure for evaluating…[these integrals]…in the general case" (p. 122).

Despite this, there are reduction formulae that can decrease the dimensionality of the integral if the correlation matrix takes one of various forms. For example, with a k-dimensional multivariate normal distribution, if there exists some $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)$ such that the correlation between the ith and jth components is given by $\rho_{ij} = \lambda_i \lambda_j$, then the integral can be reduced to a single dimension (Dunnett & Sobel, 1955). A special case of this occurs when all correlations are equal (say, ρ) as $\lambda_i = \sqrt{\rho}$ for all i satisfies this condition. This is the case here, as all interdetector correlations are given by $\rho_i^{-|+}$ (Equation (5.31)); this allows simplification of Equation (5.21) to

$$f(\boldsymbol{\omega}_i^-|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma}) = \int_{-\infty}^{+\infty} \phi(u) \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - u\sqrt{\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}}\right) \mathrm{d}u; \tag{5.32}$$

see Kotz et al. (2000, pp. 133–134). Here h_{ij} is the standardised difference between the cutoff value c and the expected conditional received signal strength; that is,

$$h_{ij} = \frac{c - \mathrm{E}(y_{ij}^{*-}|\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i)}{\sqrt{(\boldsymbol{\Sigma}_i^{-|+})_{jj}}}.$$

The second term in the numerator is the jth element in the vector given by Equation (5.28), while the denominator is a diagonal element of the variance-covariance matrix given by Equation (5.29).

Computation

The astute reader will notice that the likelihood as originally specified (Equation (5.12)) contained a one-dimensional integral, and—following substantial derivation—an alternative

specification of the joint density $f(\boldsymbol{\omega}_i, \boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\gamma})$ (Equation (5.19), with its two terms given in Equations (5.20) and (5.32), respectively) still contains a one-dimensional integral. This respecification is therefore ultimately unlikely to make a great difference in computational efficiency. This work was pursued as it was hoped that a closed-form expression for the joint density would be obtainable, but this does not appear to be the case. This is largely a result of the multivariate normal CDF not being available in closed form, and so its evaluation requires a numerical approximation. It does, however, come with some small advantages over an implementation of the likelihood as displayed in Equation (5.12) (see Section 5.5.4) and so `admbsecr` makes use of the derivation directly above.

Numerical approximation of the integral in Equation (5.32) could again make use of the so-called rectangle rule (as per integration over animal locations—Equation (2.15)—and over cue directions—Equation (5.9)); however, *Gaussian quadrature rules* are likely to give a better approximation to the integral (using the same number of quadrature points) given that the integrand is sufficiently smooth and can be written in a particular form. Specifically, *Gauss-Hermite quadrature* approximates integrals of the form

$$\int_{-\infty}^{+\infty} e^{-y^2} f(y)\,\mathrm{d}y. \tag{5.33}$$

This is particularly attractive here as (i) the integrand of Equation (5.32) can be written in this way (see below), and (ii) the domain of integration in Equation (5.32) is also $\mathbb{R}$ (cf. integration over animal locations and cue directions, which occur over $\mathcal{A}$ and $[0, 2\pi)$, respectively). Specific to the latter, while the rectangle rule could be used here, finite limits must be set on the integral. Setting these limits too close together causes inaccuracy if the integrand is not negligible outwith these, while setting them too far apart causes computational inefficiency due to the evaluation of the integrand at points in the domain at which the function is negligible in size.

Here the integral in Equation (5.32) is redefined so that it is in the form required for Gauss-Hermite quadrature (Equation (5.33)). Notation previously reserved for SECR variables, functions, and parameters is momentarily ignored. In Equation (5.32) the integrand is

$$f_u(u) = \phi(u) \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - u\sqrt{\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}}\right)$$

$$= \frac{1}{2\pi} \exp\left(\frac{-u^2}{2}\right) \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - u\sqrt{\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}}\right), \quad -\infty < u < +\infty.$$

Setting $y = u/\sqrt{2}$,

$$u = v(y) = y\sqrt{2}$$
$$v'(y) = \sqrt{2},$$

and so the change-of-variable technique gives

$$\begin{aligned} f_y(y) &= f_u(v(y))|v'(y)| \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y\sqrt{2})^2}{2}\right) \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - (y\sqrt{2})\sqrt{\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}}\right) |\sqrt{2}|, \quad -\infty < y < +\infty \\ &= \frac{1}{\sqrt{\pi}} \exp(-y^2) \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - y\sqrt{2\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}}\right), \quad -\infty < y < +\infty. \end{aligned}$$

Therefore

$$\begin{aligned} \int_{-\infty}^{+\infty} f_u(u)\,\mathrm{d}u &= \int_{-\infty}^{+\infty} f_y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} e^{-y^2} g(y)\,\mathrm{d}y \end{aligned} \tag{5.34}$$

where

$$g(y) = \frac{1}{\sqrt{\pi}} \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - y\sqrt{2\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}}\right).$$

Equation (5.34) is now in the form required for Gauss-Hermite quadrature. This gives

$$\int_{-\infty}^{+\infty} e^{-y^2} g(y)\,\mathrm{d}y \approx \sum_{i=1}^{n_q} w_i\, g(x_i), \tag{5.35}$$

where n_q is the number of quadrature points desired, x_i is the ith root of the nth physicists' Hermite polynomial,

$$H_n(x) = (-1)^n e^{x^2} \frac{\mathrm{d}n}{\mathrm{d}x^n} e^{-x^2},$$

and w_i is a weight given by

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_i)]^2};$$

see Koornwinder, Roderick, Koekoek, and Swarttouw (2010).

To summarise, estimation of $\boldsymbol{\theta}$ for a model incorporating source signal strength heterogeneity can be achieved via maximisation of Equation (5.11), using Equation (5.19) for calculation of $f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i; \boldsymbol{\gamma})$. The first term in this is given by Equation (5.20), and provided in full in Equation (5.18), while the second—Equation (5.32)—can be approximated by Equation (5.35), above.

5.3 Implementation in admbsecr

Simulation and model fitting for both CD and CSH are supported in admbsecr. The parameters β_2 (for cue directionality) and σ_v (for source strength heterogeneity) are named beta2.ss and sigma.beta0.ss, respectively.

To simulate data under each model it is only necessary to include the relevant parameter in the list used as the pars argument to sim.capt (see Section 2.3.5). For example[2],

```
dir.capt <- sim.capt(traps = traps, mask = mask, infotypes = "toa",
                     pars = list(D = 750, b0.ss = 160, b1.ss = 3,
                         b2.ss = 10, sigma.ss = 10, sigma.toa = 0.002),
                     detfn = "ss", ss.opts = list(cutoff = 130))
het.capt <- sim.capt(traps = traps, mask = mask,
                     pars = list(D = 750, b0.ss = 160, b1.ss = 3,
                         sigma.b0.ss = 4, sigma.ss = 4),
                     detfn = "ss", ss.opts = list(cutoff = 130))
```

The traps and mask objects used here are those that come with the example data (i.e., example$traps and example$mask, respectively.

Estimation of parameters for each situation simply requires setting the component directional or het.source, respectively, to TRUE in the list used as the ss.opts argument to admbsecr:

[2] Running set.seed(4321) prior to executing this code will result in the same capt objects used throughout this section.

```
dir.fit <- admbsecr(capt = dir.capt, traps = traps, mask = mask,
                    ss.opts = list(cutoff = 130, directional = TRUE))
het.fit <- admbsecr(capt = het.capt, traps = traps, mask = mask,
                    ss.opts = list(cutoff = 130, het.source = TRUE))
```

Many of the utility functions will then behave as expected; for example, `summary()` will show parameter estimates and standard errors:

```
summary(dir.fit)

# Detection function: Signal strength
# Information types: Signal strengths, Times of arrival
#
#  Parameters:
#             Estimate Std. Error
# D         1.0361e+03   225.1300
# b0.ss     1.6625e+02     4.9716
# b1.ss     4.9091e+00     0.8807
# b2.ss     9.6199e+00     3.7855
# sigma.ss  1.1718e+01     2.1324
# sigma.toa 2.0323e-03     0.0007
# ---
# esa       2.7026e-02     0.0029

summary(het.fit)

# Detection function: Signal strength
# Information types: Signal strengths
#
#  Parameters:
#               Estimate Std. Error
# D           723.120539   121.1200
# b0.ss       162.124360     1.8764
# b1.ss         3.224458     0.1553
# sigma.b0.ss   1.755212     2.4485
# sigma.ss      3.619098     0.3691
# ---
# esa           0.067762     0.0059
```

However, at the time of writing, some functions have not yet been implemented for CD or CSH models—specifically `locations()`, `show.detfn()`, and `show.detsurf()`. This is because it is necessary to integrate over an additional latent variable for each individual in order to calculate marginal detection probabilities, and the current function definitions do not yet have this capability.

For models incorporating CSH the Hermite polynomial roots and their respective weights (x_i and w_i in Equation (5.35), respectively) are obtained from the `gaussHermiteData()` function exported by the `fastGHQuad` R package (Blocker, 2014). These are then passed to ADMB via the `.dat` file. They are used as arguments to the C++ function `pmvn()` (defined in the file `ADMB/src/densfuns.cpp` within the main directory of the `admbsecr` package) which calculates the multivariate normal CDF using Gauss-Hermite quadrature within the likelihood definition found in the `secr.tpl` template file. As an alternative, calculation using the rectangle rule is also possible. This can be used by setting the `het.source.method` component of `ss.opts` to `"rect"`; in this case the limits of integration are set at ± 5, and functionality for these to be determined by the user does not presently exist. The default is `"GH"`, for Gauss-Hermite quadrature.

The number of quadrature points for both directional cue and source strength heterogeneity models can also be set. This is achieved using the `n.dir.quadpoints` and `n.het.source.quadpoints` components of `ss.opts`. These default to fifteen and eight, respectively—although increasing these is recommended in situations with strong cue directionality (i.e., a large β_2 parameter) and high intercue strength variation (i.e., a large σ_v parameter). Increasing the number of quadrature points reduces the error in the integral approximations (Equations (5.9) and (5.35)) at the cost of additional computing time.

5.4 Simulation studies

Two simulation studies were carried out: one for each of the CD and the CSH methods described in Section 5.2.

The directional cue simulation study used a detector layout identical to that of the `example` data set (see Figure 2.3) and parameter values $D = 750$, $\beta_0 = 160$, $\beta_1 = 3$, $\beta_2 = 20$, $\sigma_s = 10$, and $\sigma_t = 0.002$. In total 500 data sets were generated using `sim.capt()`, and two models were fitted to each: one that accounted for cue directionality, and one that did not. Boxplots of all parameters estimated from the simulated data using these methods are shown in Figure 5.3.

The source strength heterogeneity simulation study used a detector layout comprising

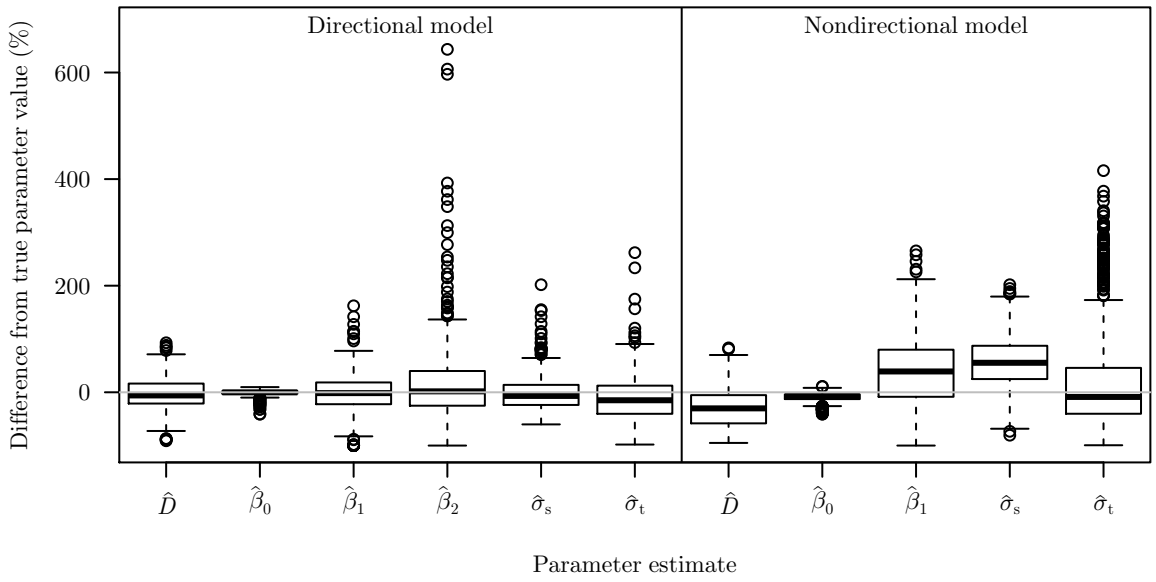


Figure 5.3 *Boxplots showing point estimates for all parameters from both the directional and nondirectional models. These are given as percentage differences from the true parameter value from which the data were generated.*

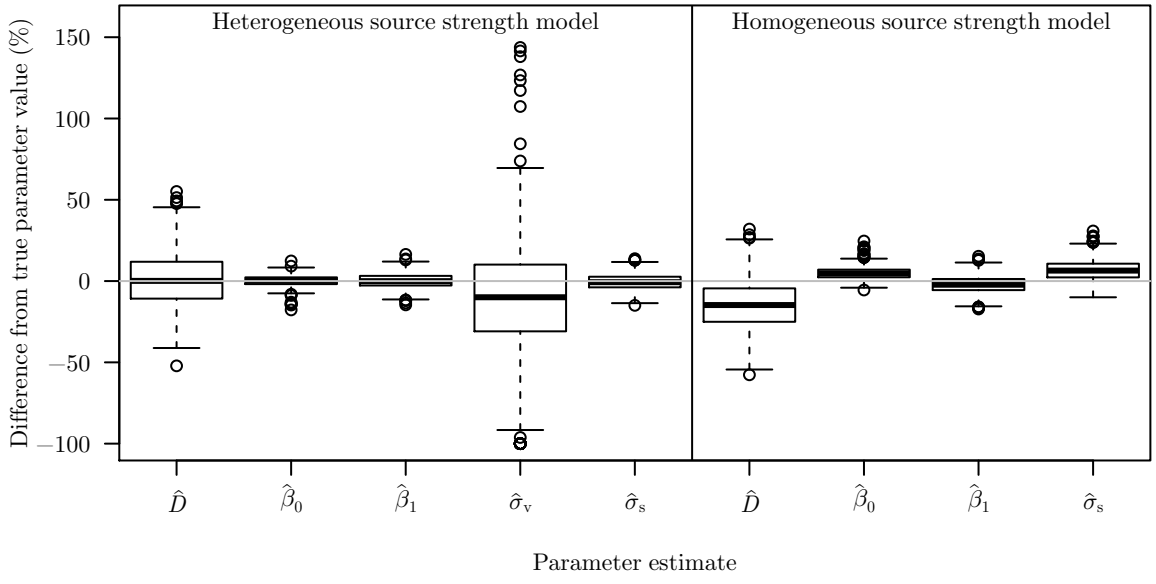


Figure 5.4 *Boxplots showing point estimates for all parameters from models that do and do not incorporate CSH. These are given as percentage differences from the true parameter value from which the data were generated.*

Table 5.1 *Estimated biases, CVs, and RMSEs of estimators for models that both do ('Yes' columns) and do not ('No' columns) account for CD (top section) and CSH (bottom section). All values are given as percentages of their underlying values.*

	Bias (%)		CV (%)		RMSE (%)	
Parameter	Yes	No	Yes	No	Yes	No
D	−1.78	−30.72	30.93	41.11	30.98	51.32
β_0	−0.95	−10.42	6.25	10.73	6.33	14.96
β_1	−2.13	32.41	36.09	79.67	36.15	86.01
β_2	20.20	—	85.26	—	87.62	—
σ_s	−1.38	57.27	34.88	51.83	34.90	77.24
σ_t	−11.08	25.62	45.83	106.25	47.15	109.30
D	0.89	−14.42	16.07	14.90	16.09	20.74
β_0	0.15	4.91	3.25	3.82	3.25	6.22
β_1	0.21	−2.09	4.57	5.07	4.57	5.48
σ_v	−10.89	—	34.79	—	36.45	—
σ_s	−0.66	6.63	4.86	6.22	4.91	9.09

a four-by-four grid of detectors with each row and column separated by one distance unit. Parameter values were $D = 7500$, $\beta_0 = 50$, $\beta_1 = 7.5$, $\sigma_v = 4$, and $\sigma_s = 4$. In total 1 000 data sets were generated using `sim.capt()`, and again two models were fitted to each: one that accounted for cue strength heterogeneity, and one that did not. Boxplots of all parameters estimated from the simulated data using these methods are shown in Figure 5.4.

Estimator biases, precisions, and MSEs estimated from both simulation studies are shown in Table 5.1. Density estimators have negligible bias for models that account for cue directionality or source strength heterogeneity; however, when these processes remain unmodelled the corresponding estimators show strong negative bias.

5.5 Discussion

The methods described in this chapter appropriately account for two mechanisms that can induce dependence between detections across a detector array on acoustic surveys, giving rise to animal density estimators with negligible bias (Figures 5.3 and 5.4, Table 5.1). The negative bias observed for estimators that do not account for these effects is unsurprising—this is a consistent finding in CR for models that fail to account for individual heterogeneity in detection probabilities, as discussed by Otis et al. (1978).

5.5.1 Mixture models for heterogeneous detectability

The use of a continuous latent variable to model heterogeneity in detectability gives rise to what is often referred to as an infinite mixture model. The use of mixture models to account for individual heterogeneity in both CR and SECR has been questioned on numerous occasions (Efford & Mowat, 2014; Link, 2003; Royle, Karanty, Gopalaswamy, & Kumar, 2009); furthermore, Pledger (2000) argues specifically that infinite mixtures (such as that considered above) can be problematic—citing both issues with model fitting, and that finite mixtures typically account for enough heterogeneity even in situations where infinite mixtures are the true data-generating process.

These concerns are understandable when information about the mixture distribution comes solely from binary capture histories; however, for acoustic surveys collecting signal strength information the observed data are far richer—they not only indicate whether or not an individual was detected, but also how detectable it was. A high received signal strength indicates that an individual was easily detectable, while one just above the threshold of detection, c, was not. These auxiliary data therefore provide the possibility to suitably estimate the distribution of source signal strengths, and, in turn, parameters of the mixing distribution for cue detectability.

This chapter has therefore shown an additional benefit of collecting auxiliary data. They not only enhance inference available from standard SECR models by improving estimator precision (Chapter 2), but here they facilitated their generalisation; without signal strength and TOA information it is unlikely that cue directions or source strength levels could have been inferred from the capture histories alone. Auxiliary data can therefore allow the relaxation of model assumptions and the estimation of further relevant parameters of interest.

5.5.2 Estimation of cue directionality and source strength variation

Interestingly, estimators for the two parameters that were specifically included to account for CD and CSH (β_2 and σ_v, respectively) performed poorly, showing both substantial bias and imprecision. Despite this, estimated biases for all other estimators were negligible—each clearly outperformed their counterpart in models that did not account for the additional CD or CSH effects (Table 5.1).

The cues' latent source strengths share many features with random effects in linear mixed effects models—for example, a marginal normal distribution and a variance component that requires estimation. The negative bias observed in the parameter σ_v is therefore perhaps unsurprising: it is well known that ML estimates of variance components in mixed effects

models are negatively biased due to a loss in degrees of freedom in the estimation of fixed effects, and this is precisely what *restricted* ML estimators (Patterson & Thompson, 1971) are intended to counter. While effort could therefore be made to provide better inference about the parameter σ_v via an alternative estimation approach, this is unlikely to make any practical difference to the estimator for animal density—the main parameter of interest—given its suitable performance here.

5.5.3 Incorporation into existing methods

The most substantial disadvantage of these models (as presented here) is that they require cue-based capture histories, but do not make any attempt to account for the dependence between the locations of cues emitted by the same individual. Although not directly applicable to real data on their own, the methodological development in this chapter provides an option to allow estimation of CD and CSH to other existing methods: both directional cue and cue source strength heterogeneity models essentially involve a respecification of the joint density $f(\boldsymbol{y}_{i\cdot 1}, \boldsymbol{\omega}_i | \boldsymbol{x}_i; \boldsymbol{\gamma})$ (where $\boldsymbol{y}_{i\cdot 1}$ specifically contains the received signal strength information corresponding to the ith detected cue) in such a way that involves marginalisation over a latent variable. Section 7.1 shows an example where this may prove to be useful.

There is thus scope for the extensions made here to be applied in addition to others—for example those described in Chapters 3 and 4—via an equivalent change to the calculation of the joint density therein. This has, however, not yet been attempted. There are potential issues with computational efficiency. For example, the simulation procedure of Chapter 3 is already computationally intensive, and this would only be compounded by increasing the dimensionality of the integral evaluated for each detected cue.

5.5.4 Alternative likelihood calculation

As discussed in Section 5.2.2, calculation of the joint density of received signal strengths and capture histories for models accounting for CSH can either be achieved via direct integration over each source strength,

$$f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i; \boldsymbol{\gamma}) = \int_{-\infty}^{+\infty} f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i, v_i; \boldsymbol{\gamma})\, f(v_i; \boldsymbol{\gamma})\, \mathrm{d}v_i, \tag{5.36}$$

Equation (5.12), or via considering the marginal joint density of the received signal strengths derived in Section 5.2.2, giving

$$\begin{aligned} f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i; \boldsymbol{\gamma}) &= f(\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+ | \boldsymbol{x}_i; \boldsymbol{\gamma})\, f(\boldsymbol{\omega}_i^- | \boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma}) \\ &= f(\boldsymbol{y}_i^{*+} | \boldsymbol{x}_i; \boldsymbol{\gamma}) \int_{-\infty}^{+\infty} \phi(u) \prod_{\{j:\omega_{ij}=0\}} \Phi\left(\frac{h_{ij} - u\sqrt{\rho_i^{-|+}}}{\sqrt{1-\rho_i^{-|+}}} \right) \mathrm{d}u, \end{aligned} \tag{5.37}$$

Equations (5.19), (5.20), and (5.32).

One advantage of the latter is that if all detectors detect the ith cue, then $\boldsymbol{\omega}_i^-$ remains empty. In this case, $f(\boldsymbol{\omega}_i^- | \boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+, \boldsymbol{x}_i; \boldsymbol{\gamma})$ disappears from Equation (5.37), resulting in the joint density

$$\begin{aligned} f(\boldsymbol{\omega}_i, \boldsymbol{y}_i | \boldsymbol{x}_i; \boldsymbol{\gamma}) &= f(\boldsymbol{\omega}_i^+, \boldsymbol{y}_i^+ | \boldsymbol{x}_i; \boldsymbol{\gamma}) \\ &= f(\boldsymbol{y}_i^{*+} | \boldsymbol{x}_i; \boldsymbol{\gamma}), \end{aligned} \tag{5.38}$$

which is therefore available in closed form as the multivariate normal PDF (Equation (5.18)) and so no integration is required; this considerably reduces computation time of likelihood components for some individuals. A naïve implementation of the former does not come with this advantage.

Of course, one may compute the joint density using Equation (5.36) when not all detectors make a detection and using the multivariate PDF (Equation (5.18)) when they do, thus gaining the same advantage. Nevertheless, some results shown in the alternative derivation remain relevant—in particular, the specifications of the inverse and determinant of the variance-covariance matrix $\boldsymbol{\Sigma}_i$ (Equations (5.25), (5.26), and (5.27)); these circumvent any computationally expensive means of their calculation within Equation (5.18).

Furthermore, recall that optimisation of the log-likelihood is carried out with the use of AD; thus it is not only the values of $\boldsymbol{\Sigma}_i^{-1}$ and $|\boldsymbol{\Sigma}_i|$ that require computation, but also their contributions to the partial derivatives of the log-likelihood with respect to the model parameters. The specifications of $\boldsymbol{\Sigma}_i^{-1}$ and $|\boldsymbol{\Sigma}_i|$ provided in Section 5.2.2 minimise the number of binary operations required for their calculation, therefore reducing not only the computation time of the log-likelihood function, but also that of its partial derivatives.

Finally, direct approximation of the above integral (and, indeed, the outer integral over individuals' latent locations) in order to compute the likelihood function could be avoided in its entirety by taking a Bayesian approach to parameter estimation; Markov chain Monte Carlo (MCMC) algorithms instead allow inference by sampling values of these

latent variables (i.e., the cue directions, u_i, and source strengths, v_i here). Bayesian methods often perform well for models that involve levels of latent variables—as is the case here—and may prove to be particularly useful for SECR models that are looking to incorporate the methods described in this chapter, potentially in addition to further complex effects (e.g., see the discussion in Section 7.1).

5.5.5 Further extensions

A variety of straightforward generalisations to the methods described in Section 5.2 are possible.

First, the methods described here have assumed the availability of signal strength information; nevertheless, they can be extended to situations without any auxiliary data. For example, models in which detection function parameters vary across individuals (as per β_0 in Section 5.2.2) can be considered in the absence of auxiliary data. However, the use of such infinite mixture models to account for heterogeneous detectability may then be considered questionable (as per the discussion in Section 5.5.1). Likewise, one may consider the use of any other detection function within a CD model, whereby its scale parameter (e.g., σ for either the halfnormal or hazard-rate detection functions) depends on the cue's direction as β_1 did in Section 5.2.1. This was not considered here simply because it is unlikely that $\boldsymbol{\Omega}$ alone holds enough information about cue directions to make implementation of such a model worthwhile. An exception to this is if the detector array comprises a large, dense grid of detectors and so—conditional on a cue's location—its detection pattern provides enough information from which its direction can be inferred.

Second, the assumed relationship between directionality and detectability was assumed to be a cosine function (Equation (5.2), Figure 5.2). For many species this may not be reasonable, and the literature may suggest the use of an alternative. For example, bursts of clicks produced by the sperm whale *Physeter macrocephalus* have been found to conform to the 'bent horn' hypothesis (Møhl, Wahlberg, Madsen, Miller, & Surlykke, 2000), whereby each cue comprises a focussed, forward-directed beam and a broader, backwards-directed beam, along with an additional omnidirectional component (Zimmer, Tyack, Johnson, & Madsen, 2005). This implies that an appropriate relationship between the bearing discrepancy and $l_j(\boldsymbol{x}, u; \boldsymbol{\gamma})$ should have two minima: one at a bearing discrepancy of 0, and another at $\pm\pi$. Given the specification of an appropriate function, such an extension is straightforward under the framework presented here.

Finally, in some cases CD may not be caused by the nature of the call production process itself, but instead some environmental variable. One obvious example is that of a prevailing

wind blowing across the survey region; detectors may be more likely to make a detection if they are downwind from a cue-producing individual. In this case cue directionality is the same for all individuals, and is directly observable. The likelihood in Equation (5.12) can be altered to account for this via substitution of the observed wind direction for v_i and removal of the inner integral sign and differential—there is no need to carry out integration over cue directions once they are observed.

5.5.6 Concluding remarks

Dependence between detections was induced by fairly specific mechanisms for the models described here. There is scope for the development of more general methodology—for example, in which distance-based correlations between detection probabilities across a detector array are estimated, thus allowing for a second layer of dependence to be modelled beyond what is already accounted for by the latent animal locations alone.

Specifically, further generalisation is required for their application to cue-based capture history data in order to obtain appropriate animal density estimates (Section 5.5.3). Perhaps the most comprehensive way for this to occur is within a cue-based model that accounts for individual identities and call rates—such as those described in both Sections 3.6.6 and 4.6.5. Further details on how this may be achieved are presented in Section 7.1.

Chapter 6

Uncertain identification in surveys of wildlife populations

6.1 Introduction

Data collection during surveys of wildlife populations routinely involves detections, and subsequent redetections, of the same animals. A record is often kept of which detections correspond to the same individual, and identifying redetections as such depends on the method of detection. If individuals are physically captured tags containing unique identifiers can often be affixed to individuals the first time they are caught.

With the continued technological advancement in the equipment used for wildlife monitoring, detection of individuals is beginning to move away from physical capture towards more passive means—such as monitoring via acoustic detectors, cameras, and genetic detection from hair snags or scat. Such approaches allow for the collection of data that was not previously obtainable; however, as surveyors further distance themselves from the animals they monitor, there is greater potential for uncertainty in animal identification. For example, recognition of individuals from images captured by high-definition video can be achieved from natural markings or scars; however, this may only be possible if the redetection occurs from the same viewing angle as the initial detection. Even genetic identification becomes uncertain in a variety of circumstances, for example the presence of allelic dropout (Gagneux, Boesch, & Woodruff, 1997; Taberlet et al., 1996). While methods for automatic recognition of individuals from their calls (i.e., without resorting to a manual allocation of identities to detections by a surveyor—potentially a tedious and time-consuming task) are currently being developed (e.g., Foote, Palazzi, & Mennill, 2013; Laiolo, Vögeli, Serrano, & Tella, 2007; Petrusková, Pišvejcová, Kinštová, Brinke, & Petrusek, 2016; Xia, Huang, Wei,

Nie, & Zhang, 2001), these may not provide reliable enough identities for variety of species. Indeed, state-of-the-art machine learning approaches for identifying bird species (let alone individuals from the same species) still do so with some degree of uncertainty (see Stowell & Plumbley, 2014). Even surveys that do involve the physical capture and tagging of animals may be subject to uncertain identification, as some individuals may lose their tags.

A small fraction of statistical approaches used to estimate parameters of interest from these data does not require any information about the identity of detected individuals, and instead replace this information with (potentially restrictive) assumptions. *Random encounter models* (REMs; Rowcliffe, Field, Turvey, & Carbone, 2008) are one such example, whereby animals are assumed to behave akin to particles of an ideal gas (Hutchinson & Waser, 2007). To fulfil this assumption individuals must move in straight lines and must not interact with one another. Applications of REMs are almost entirely confined to data collected by camera traps—though a recent generalisation also allows for their use with acoustic detection data (Lucas, Moorcroft, Freeman, Rowcliffe, & Jones, 2015). This approach is not appropriate for species where individuals move a negligible distance over the course of a survey (e.g., *A. lightfooti*; see Section 3.5.1), or move in such a way that deviates markedly from particles of an ideal gas. Furthermore, when partial information about identity *is* available, approaches like REMs neglect its use entirely.

Almost all other developed methods require complete information about the identity of the individual that corresponds to each detection; that is, it must be known with certainty whether or not any given two detections are detections of the same animal. This holds for most CR methods: records of when (and where, for SECR) each individual was detected must be observed. The methodology presented in all preceding chapters of this thesis assumed that it was known whether or not any two detections by any two detectors were of the same cue. In Chapter 4 it was further assumed that it was known whether or not any two detected cues were emitted by the same individual.

The development of statistical methodology that is capable of estimating parameters of interest from data where identification is uncertain—but some information is available—is therefore important. Recently progress has been made for specific causes of misidentification in CR surveys—for example, accounting for allelic dropout affecting genetic identification (Wright et al., 2009) or adjusting for photographic identification errors in visual surveys (Morrison, Yoshizaki, Nichols, & Bolger, 2011). More generally, model M_t from Otis et al. (1978; see Section 1.1.2) has been extended to allow for misidentification of individuals across sampling occasions (Hiby et al., 2012; Link, Yoshizaki, Bailey, & Pollock, 2010; Schofield & Bonner, 2015; Tancredi, Auger-Méthé, Marcoux, & Liseo, 2013; Vale, Fewster,

Carroll, & Patenaude, 2014).

This chapter broaches a problem that is more general still: the analysis of detection data for which no attempt is ever made to identify individuals. Therefore, unlike the methods outlined above, there is no 'misidentification' to speak of—though noisy information about animal identities is contained within the data. A new general approach is developed for the estimation of ecological parameters of interest from such data, giving rise to a class of *trace-contrast* (TC) *models*. The formulation of these models is based on the simple idea that data (a *trace*) can be collected from each detection, and that traces from two detections of the same individual are likely to be more similar (or *proximal*) than detections from each of two different, randomly selected individuals. Estimation is based on modifications to an approach proposed by Tanaka, Ogata, and Stoyan (2008) for the analysis of spatial data that are assumed to be a realisation of a Neyman Scott point process (NSPP); see Neyman and Scott (1952).

The issue of individual identification has been referred to throughout the previous chapters of this thesis. For example, the methods described in Chapter 4 assumed identities were observed, while those of Chapter 3 did not. The latter instead required the collection of independent cue-rate data. The intention is that the future development of TC models will allow for direct estimation of animal density from cue-based SECR surveys that do not record identities (i.e., without the collection of these additional data). Conceptually, assuming that individuals are stationary, detected cues produced by the same animal should have similar (or proximal) spatial detection patterns (i.e., capture histories) and sets of observed auxiliary information that are also proximal (in that they are consistent with cues produced at similar locations). However, deriving appropriate TC model estimators from such SECR surveys is complex, and will not be considered within this thesis. Instead, here the development of TC methods is accompanied by a simpler motivating example: *two-plane surveys* designed to estimate cetacean density.

The remainder of this introduction outlines NSPPs (Section 6.1.1); summarises the estimation approach due to Tanaka et al. (2008; Section 6.1.2); and describes two-plane survey protocol, the type of data collected, and existing estimation methods (Section 6.1.3). An overview of the chapter follows (Section 6.1.4).

As this material is a departure from the development and application of SECR methodology, the notational consistencies used previously are ignored in this chapter. A summary of notation specific to this chapter can be found appended to the main notation list. Here uppercase is used to denote a random variable and lowercase denotes a fixed value of some kind (e.g., an observed value). Functions $f(\cdot)$ and $F(\cdot)$ are still used to denote PDFs/PMFs

and CDFs, respectively. Both $E(\cdot)$ and $V(\cdot)$ take a random variable's parameter vector as the only argument and return the expectation and variance, respectively. The function $C(\cdot)$ returns the covariance between two random variables. Functions $f(\cdot)$, $F(\cdot)$, $E(\cdot)$, $V(\cdot)$, and $C(\cdot)$ are all immediately followed by a lowercase subscript to indicate the random variable(s) they are associated with. Thus, $f_x(x)$ is the probability density/mass of the random variable X at some fixed value x. For example, if X is Gaussian and $\boldsymbol{\theta} = (\mu, \sigma)$ contains its expectation and standard deviation, then $E_x(\boldsymbol{\theta}) = \mu$ and $V_x(\boldsymbol{\theta}) = \sigma^2$. Further notation is provided over the following sections.

6.1.1 Neyman-Scott point processes

NSPPs give rise to clustered point pattern data. In this section they are described as per the definition of Illian, Penttinen, Stoyan, and Stoyan (2008, p. 374, though with differing notation) in two-dimensional space, but they can be extended to an arbitrary number of dimensions.

Clustering is induced by *parent points*, locations of which are a realisation of some point process with homogeneous intensity D. While the locations of parents remain unobserved, they each spawn a number of observable *children points*. The numbers of children generated by parents are typically assumed to be IID from some discrete distribution characterised by parameters $\boldsymbol{\psi}$. Conditional on their parents' locations, children are IID in two-dimensional space according to some bivariate normal distribution, each with an expectation vector given by the location of its parent and remaining parameters $\boldsymbol{\gamma}$ defining the variance-covariance matrix. The vector of all parameters is given by $\boldsymbol{\theta} = (D, \boldsymbol{\psi}, \boldsymbol{\gamma})$.

More specifically, it is common to assume that (i) parent point locations are generated by a Poisson point process, (ii) the number of children generated by each parent is a Poisson random variable with expectation ν, and (iii) the variance-covariance matrix for the Gaussian dispersion of children around parents is isotropic, with the form $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_2$; thus, in this case, $\boldsymbol{\theta} = (D, \nu, \sigma)$. See Figure 6.1 for an example of a realisation of such an NSPP.

6.1.2 Neyman-Scott point process parameter estimation

Estimation of $\boldsymbol{\theta}$ for NSPPs is notoriously difficult (Baudin, 1981). It is not known how many parent points exist, nor where those that do exist are located. The identity of each child point's parent is also not observed, thus it is not known whether or not any two points are *siblings* (i.e., children points with the same parent).

Due to these latent processes the likelihood function for ML estimation is thought to be intractable (Baudin, 1981; Tanaka et al., 2008); its computation would require (i) summing

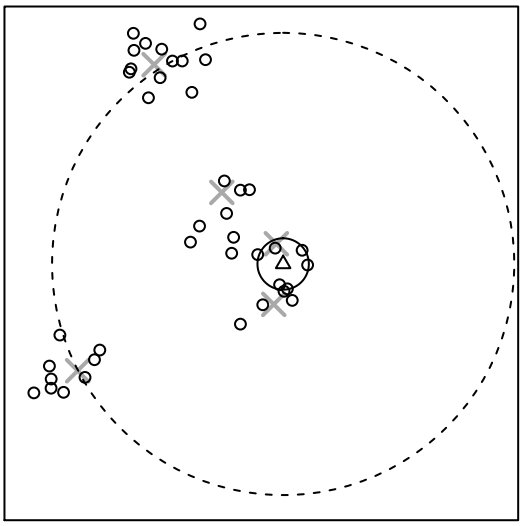


Figure 6.1 *A simulated realisation of an NSPP on the unit square. Parent points are generated by a homogeneous Poisson point process with intensity $D = 5$. The numbers of children spawned by the parents are realisations of IID Poisson random variables with expectation $\nu = 8$. The dispersion of the children around their parents is due to a bivariate normal distribution, $\mathrm{N}_2(\mathbf{0}_2, \sigma^2 \boldsymbol{I}_2)$, with $\sigma = 0.05$. As this is a simulated point pattern, parent locations are known and are given by grey crosses. Locations of child points are given by open plotting characters. Note that for these data the children are loosely aggregated into three clusters despite their being five parent points; this is due to three parents being in close proximity. If given the child locations alone, one may naïvely consider these data consistent with a NSPP with a lower parent point density, larger family sizes, and slightly more dispersion of children around their parents—say $D = 3$, $\nu = 8$, and $\sigma = 0.1$. For conceptual illustration of the PI function, the triangle represents the location of a selected child point. At short distances from this focal point (solid circle), intensity of the point pattern is high due to both nearby sibling and nonsibling points. At large distances (dashed circle), intensity has fallen to a baseline rate due solely to nonsibling points.*

over all possible numbers of existing parent points, (ii) integration over all possible locations of these parent points, and (iii) summing over all possible allocations of children to the parents. The inner summand would then involve a product of probability densities due to the dispersion of children around their allocated parent. The number of combinations associated with (iii), above, is particularly prohibitive: letting m and n be the fixed number parent and child points, respectively, there are n^m possible allocations of children to parents.

A general estimation approach therefore cannot involve any allocation of children to parents. A substantial breakthrough that meets this criterion is due to Tanaka et al. (2008), who described a method that provides appropriate estimates of D, ν, and σ for the special case outlined above in two dimensions[1]: Parent locations generated by a homogeneous

[1]Tanaka et al. (2008) recognised that generalisation to NSPPs in any dimension is possible, though details were not explicitly provided.

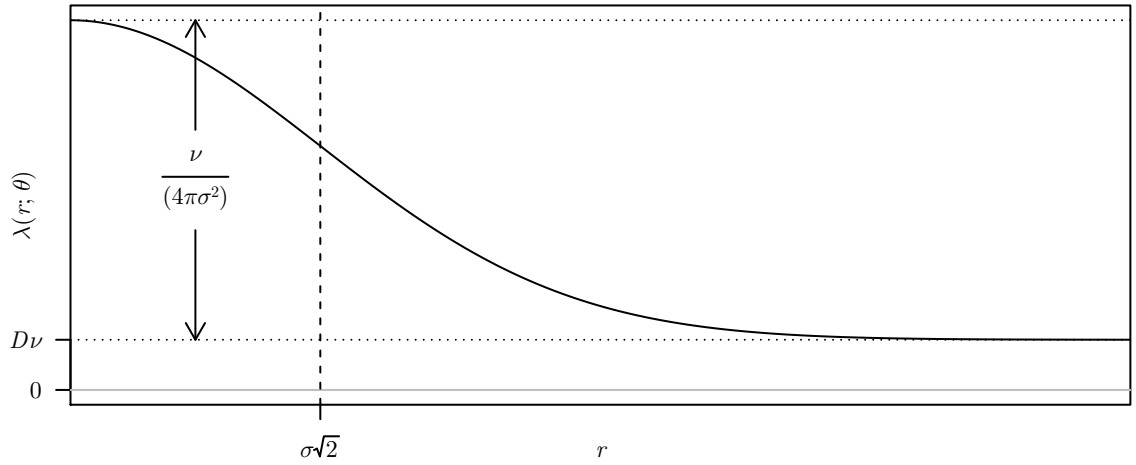


Figure 6.2 *The functional form of the PI derived by Tanaka et al. (2008). Three combinations of the parameters affect different features of the function: (i) the horizontal asymptote ($D\nu$), (ii) the difference between the horizontal asymptote and the y-intercept ($\nu/(4\pi\sigma^2)$), and (iii) the point of inflection of the Gaussian term ($\sigma\sqrt{2}$).*

Poisson point process, the numbers of children spawned by the parents being IID Poisson random variables[2], and dispersion of children around parents due to a bivariate normal distribution with a variance-covariance matrix of any (positive-definite) form[3]. The remainder of this section outlines the method, as presented by Tanaka et al. (2008).

The Palm intensity function

The approach revolves around the *Palm intensity* (PI) *function*, $\lambda(r;\boldsymbol{\theta})$. This returns the expected intensity of the point process at distance r from the location of a randomly selected, observed point. As NSPPs result in clustered point patterns, the PI is high when r is small—the presence of the observed point suggests there are nearby siblings. As r increases the PI decays to an asymptote at the overall average intensity of the point process (see Figure 6.1 for a conceptual illustration, and Figure 6.2 for the functional form of the PI function itself). Under the conditions set out above the PI function is available analytically:

$$\lambda(r;\boldsymbol{\theta}) = D\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(\frac{-r^2}{4\sigma^2}\right). \tag{6.1}$$

[2]Tanaka et al. (2008) claimed that their method is appropriate for any discrete distribution; however, this is not the case. More detail is provided in Section 6.4.4.

[3]This was noted by Prokešová and Vedel Jensen (2013); Tanaka et al. (2008) only recognised that their method was appropriate with variance-covariance matrices of the form $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_2$, giving rise to an isotropic point process.

The first term gives the PI due to nonsibling points. As parent points are generated independently, this is simply the overall intensity of the point process, given by the product of parent intensity and the expected number of children spawned by each parent. The second term, a Gaussian function, provides the PI due to sibling points. The sum of the two therefore gives the overall PI at distance r.

As the PI function (Equation (6.1)) depends on the model parameters, estimation of $\boldsymbol{\theta}$ can be carried out by determining the parameter values that give rise to a PI function that provides the 'best' (by some measure—see below) fit to the data at hand.

Approximation of the difference process

Tanaka et al. (2008) proposed determining the 'best' fit via the *difference process*. Let $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ be a realisation of a NSPP, where $\boldsymbol{x}_i = (x_{i1}, x_{i2})$ defines the Cartesian coordinates of the ith observed point. The realisation of the difference process is found by calculating all pairwise differences between the observed points (Figures 6.3(i)–(v)). Let $\boldsymbol{d}_{ij} = (x_{j1} - x_{i1}, x_{j2} - x_{i2})$ be the difference between the ith and jth points, and $\boldsymbol{d}$ denote a point in the associated sample space, $\boldsymbol{\Omega}_d$. Note that differences above some arbitrary threshold distance from the origin, t, are truncated to avoid edge effects, and so $\boldsymbol{\Omega}_d = \{\boldsymbol{d} : \boldsymbol{d} \in \mathbb{R}^2, \|\boldsymbol{d}\| < t\}$.

As the PI function, $\lambda(r; \boldsymbol{\theta})$, gives the expected point intensity at distance r from the location of a randomly selected point, the expected intensity surface of one set of differences from a particular focal point is given directly by $\lambda(\|\boldsymbol{d}\|\,; \boldsymbol{\theta})$. That is, the intensity is given by the PI function, using the radial distance from the origin as its argument. As the difference pattern (Figure 6.3(v)) is generated via superposition of n sets of differences (Figure 6.3(iv)), the expected intensity of the difference process is given by $n\,\lambda(\|\boldsymbol{d}\|\,; \boldsymbol{\theta})$ (Figure 6.3(vi)). This only holds in the absence of edge effects—to avoid these truncation of differences must occur at some distance less than the maximum possible distance between two points in the original process, and all differences must be calculated subject to periodic boundary conditions (PBCs; Figure 6.3(iii)).

Thus, the expected intensity surface of the difference process is an analytically tractable function that depends on the model parameters. Tanaka et al. (2008) suggested the fitting of an inhomogeneous Poisson point process to the difference pattern using this known functional form. The likelihood for a Poisson point process is well known, and so estimation of $\boldsymbol{\theta}$ can be carried out via its maximisation. This objective function was coined the *Palm*

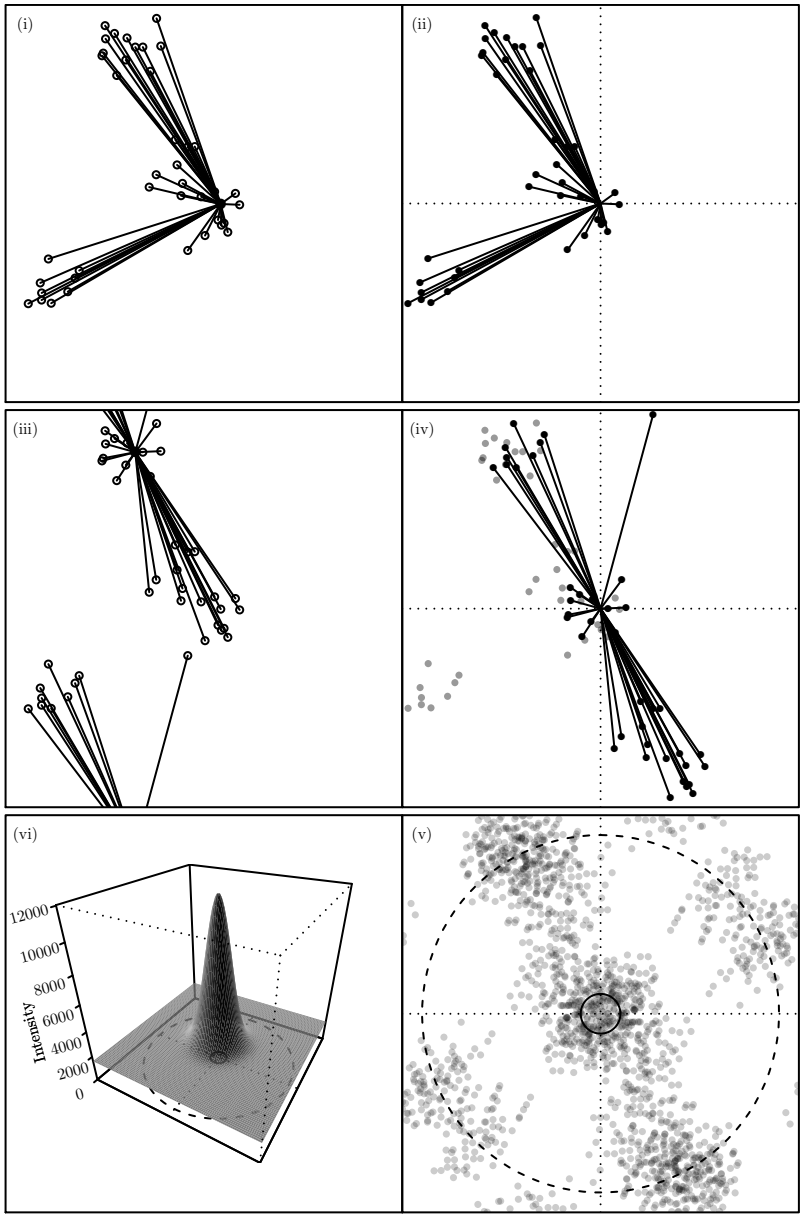


Figure 6.3 *An illustration the difference process: (i) A focal point is chosen, and differences between this point and all others are found. (ii) These differences are saved. (iii) Another focal point is chosen, and Step (i) is repeated; note the use of PBCs. (iv) The differences due to the new focal point are superimposed onto those from the initial focal point. (v) Steps (i)–(iv) are repeated for all points in the original point pattern, giving rise to the difference pattern. Similarly to Figure 6.1, at short distances from the origin (solid circle) intensity of the difference pattern is high due to differences between both sibling and nonsibling points. At large distances (dashed circle) intensity has fallen to a baseline rate due solely to differences between nonsibling points. The intensity of the difference process can be approximated by a surface that is radially proportional to the Palm intensity, Figure (vi)—here calculated using the parameter values used to generate the data.*

likelihood, and is given by

$$L(\boldsymbol{\theta}; \boldsymbol{r}) = \left(\prod_{\{i,j:i \neq j, r_{ij} < t\}} n\, \lambda(r_{ij}; \boldsymbol{\theta}) \right) \times \exp\left(-n \int_0^t \lambda(r; \boldsymbol{\theta})\, 2\pi r \, \mathrm{d}r \right), \tag{6.2}$$

where $r_{ij} = \|\boldsymbol{d}_{ij}\|$ is the radial distance from the origin to the difference between the ith and jth points (or, equivalently, the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$). The first term is the product of the expected difference process intensities evaluated at the observed locations of the points in the difference pattern within distance t of the origin. The second term is the volume beneath the intensity surface (Figure 6.3(vi)), giving the expected number of points in the difference pattern within distance t of the origin. This is the volume of revolution of $n\, \lambda(r; \boldsymbol{\theta})$ around the intensity axis, and is available in closed form:

$$\int_0^t \lambda(r; \boldsymbol{\theta})\, 2\pi r \, \mathrm{d}r = \nu \left[\pi D t^2 + 1 - \exp\left(\frac{-t^2}{4\sigma^2} \right) \right]. \tag{6.3}$$

Figure 6.3(v) betrays an inherent weakness of this approach: though an inhomogeneous Poisson point process is used to model the difference process, this is clearly a misspecification. The modelled intensity surface (Figure 6.3(vi)) only accounts for the cluster that occurs at the origin, but not those that are generated around the periphery due to abundances of differences being taken between points from the same two groups. Fitting a Poisson point process allows for straightforward estimation of $\boldsymbol{\theta}$, but is an *ad hoc* modelling decision. Tanaka et al. (2008) nevertheless showed empirically (via simulation) that the method gives rise to estimators with negligible bias. General consistency and asymptotic normality of these estimators has since been shown (Prokešová & Vedel Jensen, 2013), providing theoretical justification.

The term 'Palm-likelihood' is therefore something of a misnomer—it is a likelihood, but not for the model that generated the data. Treating it as such (for purposes other than point estimation) is likely to be dangerous, though Tanaka et al. (2008) did not broach this issue. Problems regarding model selection and variance estimation are currently unresolved.

6.1.3 Two-plane surveys for the collection of cetacean detection data

Aerial surveys can be used to collect cetacean detection data via visual observation of animals on the surface, and provide advantages over boat surveys, including (i) they cover much larger survey areas in a shorter period of time (Slooten, Dawson, & Rayment, 2004); (ii) they are considerably cheaper; and (iii) boats often attract or repel animals (e.g., Slooten

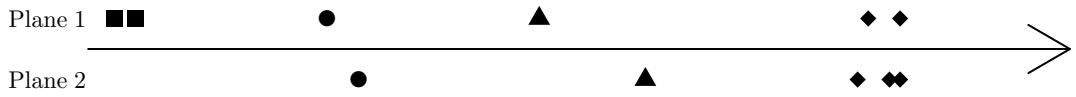


Figure 6.4 *Example data collected from a two-plane survey. Point locations along the arrow correspond to detection locations along the transect by one of the two planes. Note the four groups of points, defined by the plotting symbols. Identities associated with the square points are unambiguous—detections from the same plane cannot be of the same animal. Identities associated with the circle points are ambiguous and there are two possible allocations—these points could represent two detections of the same individual, or one detection each of two different individuals. Identities associated with the triangle points are unambiguous—the distance between the detections is too great for these to correspond to the same individual. Identities associated with the diamond points are ambiguous—between three and five individuals have been detected, and there are thirteen different possible allocations within this group alone. Therefore, along with the ambiguous circle group, there are a total of 26 different possible allocations in this small example data set.*

& Dawson, 1994), causing bias that must be corrected for (Dawson, Slooten, DuFresne, Wade, & Clement, 2004).

Submerged individuals cannot be detected from a plane. While a single plane is capable of detecting some individuals, the resulting data do not allow for estimation of overall abundance—the proportion of individuals that have been missed is not estimable. A solution is to fly two planes in tandem. If they are far enough apart, detections of a particular group by the planes can be considered independent; in this case, if animals can be identified from sightings (i.e., it is known whether or not two detections are of the same individual), then the resulting data can be analysed using simple CR methods with each plane considered a sampling occasion. This allows for estimation of detection probability, and therefore abundance. If the planes are closer detection is not independent: an individual on the surface when the first plane flies over is more likely to be at the surface on the passing of the second plane. Methods that model this dependence in various ways exist for situations where animal identities are available (Borchers & Langrock, 2015; Okamura, Minamikawa, Skaug, & Kishiro, 2012; Schweder, Skaug, Langaas, & Dimakos, 1999), and those where they are not (Borchers, Nightingale, & Fewster, in prep.; Hiby & Lovell, 1998).

In the latter case, the data collected simply comprise the locations of the detections made (in terms of distance along the transect) and the plane that made the detection (Figure 6.4). Both methods (Borchers et al., in prep.; Hiby & Lovell, 1998) involve ML estimation, and evaluate the likelihood function by treating the unknown identities associated with the detections as latent variables. This approach therefore leads to a sum over all possible allocations of identities. There are some constraints, however, that make this computationally tractable: (i) two detections made by the same plane cannot be of

the same individual (as the planes travel faster than any animal); and (ii) two detections, one from each plane, cannot be of the same individual if the difference between them is larger than the distance an animal can travel during the interplane time period—see Figure 6.4. Enumerating all possible identity allocations is challenging in itself (Borchers et al., in prep., dealt with this issue via constraint programming), and calculating all likelihood contributions across these is potentially time consuming.

A special case of the methodology developed in this chapter presents a way of estimating cetacean density without any identity allocation. This allows for faster computation, and also the ability to extract estimates from data in cases where the approaches of Hiby and Lovell (1998) and Borchers et al. (in prep.) are not computationally tractable (e.g., due to large clusters of detections, resulting in too many possible allocations to deal with).

6.1.4 Chapter overview

The majority of this chapter is devoted to deriving a general formulation for TC models (Section 6.2), including an explanation of the conceptual link between NSPPs and wildlife surveys with unknown animal identification (Section 6.2.1). This involves generalisations of the method proposed by Tanaka et al. (2008) to (i) NSPPs in a space of any dimension, (ii) any distribution for the numbers of children spawned by the parents, and (iii) allow use of known sibling (or known nonsibling) relationships between some pairs of points. A software implementation of these methods—in the form of the R package `nspp`[4]—is introduced in Section 6.3. Section 6.4 contains derivations of components of the estimation process specific to a special case of the model class, suitable for fitting to data similar to that which would be collected on a two-plane cetacean survey. The resulting estimators are shown to perform well when applied to simulated data. Discussion of further generalisations and possibilities for improved estimation procedures are presented in Section 6.5.

A manuscript recently submitted to *Statistical Science* (Fewster, Stevenson, & Borchers, in press) introduces TC models and thus overlaps with some material presented in this chapter.

[4] At the time of writing this is only available from `https://github.com/b-steve/nspp`, but the intention is for this to eventually be available on CRAN

6.2 TC models

6.2.1 Linking NSPPs to uncertain identification

Consider a wildlife survey that is capable of making multiple detections of the same individual without explicitly recording animal identity. Despite this, data (hereafter, 'traces') collected from two detections of the same individual are likely to be proximal, while those collected from detections of different animals are more likely to be father apart. For example, even in the presence of allelic dropout, genetic traces from two detections of the same individual are likely to be similar. Likewise, for camera-trap data of ship rats *Rattus rattus* analysed by Fewster et al. (in press), a single approach of an individual to a camera station triggered a glut of detections in a short space of time—thus, detection times (temporal traces) were likely to be proximal for observations of the same animal.

Indeed, for two-plane cetacean surveys, two observed locations (spatial traces) of the same individual are again likely to be proximal (i.e., detections of the same individual are likely to be at similar locations). Thus, the hypothetical two-plane survey data displayed in Figure 6.4 can be considered to be realised data from a TC model—in this case, the location of a parent point can conceptually be considered as an individual's 'average' location across the course of the survey. Cetacean density is given by D. Individuals are not observed at the parent point location due to animal movement, controlled by the parameter σ; this is equivalent to modelling this movement as diffusion. If it is assumed that the distance between an observed location and its parent is Gaussian the observed detection locations can be assumed to be a realisation of a NSPP[5]. In this case, there are distinctions between data from two-plane cetacean surveys (Section 6.1.3) and the description of NSPPs provided in Section 6.1.1, in that (i) observed cetacean locations form a one-dimensional, rather than two-dimensional, point pattern[6], and (ii) two detections from the same plane cannot be of the same individual, and so some pairs are known to be children generated by different parents. Additionally, the distribution of the number of children per parent (i.e., the distribution of the number of detections of a randomly selected individual) is clearly not Poisson: with only two planes, each animal can only be detected twice, at most. The generalisations of the method due to Tanaka et al. (2008), described below, deal with these three features of the data.

[5]Though note that not all TC models involve NSPPs. For example, in the analysis described by Fewster et al. (in press) the dispersion of traces around parents is not considered to be Gaussian—thus placing the modelled point process outwith the definition of a NSPP provided by Illian et al. (2008, p. 374).

[6]Strictly, the high-definition cameras take images of a strip of ocean and locations are observed in two-dimensions. These are collapsed to one-dimensional locations for simplicity here. See Section 6.5.4 for further discussion.

6.2.2 Parameter estimation for generalised NSPPs

In this section parameter estimators for NSPPs in general d-dimensional Euclidean space are described. As per Section 6.1.2, these are available once the Palm likelihood has been derived. Functions that depend on the dimensionality of the point process without explicitly taking this as an argument appear with a (d) superscript.

Let the observation window be $\mathcal{W} \subset \mathbb{R}^d$. Only children within this window are observed, although parents of observed children could plausibly be located outwith $\mathcal{W}$. Further, let the number of children generated by the ith parent be C_i, and, generically, the number of children generated by a randomly selected parent be C. The random variables $\{C_1, C_2, \cdots, C_m\}$ are IID, each with PDF $f_c(c; \boldsymbol{\psi})$.

Expected point counts

Before describing the PI function it is necessary to derive an expression for $n^{(d)}(r; \boldsymbol{\theta})$, the expected number of children within distance r of a randomly selected child point. This can be partitioned into points with the same, $n_s^{(d)}(r; \boldsymbol{\psi}, \sigma)$, and different, $n_d^{(d)}(r; D, \boldsymbol{\psi})$, parents; that is,

$$n^{(d)}(r; \boldsymbol{\theta}) = n_d^{(d)}(r; D, \boldsymbol{\psi}) + n_s^{(d)}(r; \boldsymbol{\psi}, \sigma). \tag{6.4}$$

The average intensity of the point process is given by $D\,E_c(\boldsymbol{\psi})$: parent density, multiplied by mean number of children spawned by each. As the latent parent locations are assumed to be a realisation of a Poisson point process, the location of one parent is independent of both the abundance of other parents, as well as their locations. Therefore, focussing on single observed child, the local intensity of children that do not share a parent with this focal point remains $D\,E_c(\boldsymbol{\psi})$. Thus the expected number of nonsibling children (i.e., children with a different parent) within distance r of an observed child is given by

$$n_d^{(d)}(r; D, \boldsymbol{\psi}) = D\,E_c(\boldsymbol{\psi})\,v^{(d)}(r),$$

where

$$v^{(d)}(r) = \frac{\pi^{d/2}\,r^d}{\Gamma(d/2+1)}$$

is the volume of the d-dimensional hypersphere with radius r.

Let Y be the distance between two randomly selected siblings, and S be the number of children with the same parent (the *family size*) as a randomly selected child—including the selected point itself. The latter is subtly different from the random variable C in that C corresponds to the number of children generated by a randomly selected *parent*, while S

corresponds to the number of children generated by the parent of a randomly selected *child*. For C, all groups are equally likely to be selected—but for S, selection of a child from a larger group is more likely. The distributions of both C an S are fully characterised by the parameters $\boldsymbol{\psi}$.

The expected number of siblings owned by a randomly selected child (not counting itself) is therefore $E_s(\boldsymbol{\psi}) - 1$, and the expected proportion of siblings within distance r of a randomly selected child is some function $F_y^{(d)}(r; \boldsymbol{\theta})$. The expected number of siblings within distance r of an observed point is then given by

$$n_s^{(d)}(r; \boldsymbol{\psi}, \sigma) = [E_s(\boldsymbol{\psi}) - 1] \times F_y^{(d)}(r; \sigma). \tag{6.5}$$

Expressions for $E_s(\boldsymbol{\psi})$ and $F_y^{(d)}(r; \sigma)$ are required in order to complete full derivation of Equation (6.4).

The PMF $f_s(s; \boldsymbol{\psi})$ is required in order to obtain an expression for $E_s(\boldsymbol{\psi})$. This can be obtained by noting that

$$f_s(s; \boldsymbol{\psi}) \propto s\, f_c(s). \tag{6.6}$$

That is, the probability of a randomly selected child belonging to a group of size s is related to both the proportion of groups that are of size s, $f_c(s)$, but also the size of these groups themselves. For example, if all group sizes are equally likely, then selection of a child belonging to a group twice as large as another is twice as likely, as twice the number of children belong to groups of this size. The PMF of S can then be obtained via normalisation of Equation (6.6):

$$f_s(s; \boldsymbol{\psi}) = \frac{s\, f_c(s)}{\sum_{i=0}^{\infty} i\, f_c(i)} = \frac{s\, f_c(s)}{E_c(\boldsymbol{\psi})}.$$

Therefore,

$$\begin{aligned} E_s(\boldsymbol{\psi}) &= \sum_{i=0}^{\infty} i\, f_s(i; \boldsymbol{\psi}) \\ &= \sum_{i=0}^{\infty} i\, \frac{i\, f_c(i)}{E_c(\boldsymbol{\psi})} \\ &= \frac{1}{E_c(\boldsymbol{\psi})} \sum_{i=0}^{\infty} i^2\, f_c(i) \\ &= \frac{E_{c^2}(\boldsymbol{\psi})}{E_c(\boldsymbol{\psi})} \\ &= \frac{V_c(\boldsymbol{\psi}) + E_c(\boldsymbol{\psi})^2}{E_c(\boldsymbol{\psi})}. \end{aligned} \tag{6.7}$$

Note that for the special case $C \sim \text{Poisson}(\nu)$

$$\begin{aligned} E_s(\boldsymbol{\psi}) - 1 &= \frac{V_c(\boldsymbol{\psi}) + E_c(\boldsymbol{\psi})^2}{E_c(\boldsymbol{\psi})} - 1 \\ &= \frac{\nu + \nu^2}{\nu} - 1 \\ &= \nu = E_c(\boldsymbol{\psi}). \end{aligned} \tag{6.8}$$

In order to derive $F_y^{(d)}(r;\sigma)$, let $\boldsymbol{X}_a = (X_{a1}, \cdots, X_{ad})$ and $\boldsymbol{X}_b = (X_{b1}, \cdots, X_{bd})$ define the locations of two randomly selected siblings with parent location $\boldsymbol{p} = (p_1, \cdots, p_d)$. Recall that—for an isotropic NSPP—the nature of dispersion of children around their parents gives the result

$$\boldsymbol{X}_a, \boldsymbol{X}_b \sim \text{N}_d(\boldsymbol{p}, \sigma^2 \boldsymbol{I}_d).$$

The random variable Y is the distance between two randomly selected siblings, and so

$$Y = \|\boldsymbol{X}_b - \boldsymbol{X}_a\| = \sqrt{\sum_{i=1}^{d} (X_{bi} - X_{ai})^2}, \tag{6.9}$$

where

$$X_{ai}, X_{bi} \sim \text{N}(p_i, \sigma^2),$$

and thus

$$X_{bi} - X_{ai} \sim \text{N}(0, 2\sigma^2).$$

Now, let

$$Z_i = \frac{X_{bi} - X_{ai}}{\sqrt{2\sigma^2}},$$

resulting in $Z_i \sim \text{N}(0,1)$. This gives

$$X_{bi} - X_{ai} = Z_i \sqrt{2\sigma^2}, \tag{6.10}$$

and substitution of Equation (6.10) into Equation (6.9) provides

$$\begin{aligned} Y &= \sqrt{\sum_{i=1}^{d} \left(Z_i \sqrt{2\sigma^2} \right)^2} \\ &= \sqrt{2\sigma^2 \sum_{i=1}^{d} Z_i^2} \end{aligned}$$

$$
\begin{aligned}
&= \sigma\sqrt{2\chi^2} \\
&= \sigma\chi\sqrt{2},
\end{aligned}
$$

where $\chi^2 \sim$ chi-square(d) and $\chi = \sqrt{\chi^2} \sim$ chi(d), the latter having PDF

$$
f_\chi(x; d) = \frac{2^{1-d/2}\, x^{d-1} e^{-x^2/2}}{\Gamma(d/2)},
$$

see Johnson, Kotz, and Balakrishnan (1995) for a description of the chi distribution.

The PDF of y is then available via a straightforward application of the change-of-variable technique:

$$
\begin{aligned}
f_y^{(d)}(y; \sigma) &= f_\chi(\chi; d)\, \frac{d\chi}{dy} \\
&= f_\chi\left(\frac{y}{\sigma\sqrt{2}}; d\right) \left(\frac{d}{dy}\, \frac{y}{\sigma\sqrt{2}}\right) \\
&= \frac{2^{1-d/2}\, [y/(\sigma\sqrt{2})]^{d-1}\, \exp(-[y/(\sigma\sqrt{2})]^2/2)}{\Gamma(d/2)}\, \frac{1}{\sigma\sqrt{2}} \\
&= \frac{2^{1-d/2}\, y^{d-1}\, \exp(-y^2/(4\sigma^2))}{(\sigma\sqrt{2})^d\, \Gamma(d/2)},
\end{aligned} \tag{6.11}
$$

and the CDF is also readily available due to the transformation $y = \sigma\chi\sqrt{2}$ being strictly monotonic; this gives

$$
\begin{aligned}
F_y^{(d)}(y; \sigma) &= F_\chi\left(\frac{y}{\sigma\sqrt{2}}; d\right) \\
&= P\left(\frac{d}{2}, \frac{y^2}{4\sigma^2}\right),
\end{aligned} \tag{6.12}
$$

where $P(\cdot, \cdot)$ is the regularised gamma function[7].

Equations (6.4), and, by extension, (6.5), have now been fully derived.

The PI function

Naturally, $n^{(d)}(r; \boldsymbol{\theta})$ increases with r: increasing the radius of the d-dimensional hypersphere centred on an observed child will also increase the expected number of other children that lie within. The first derivative of $n^{(d)}(r; \boldsymbol{\theta})$ with respect to r gives the expected increase in the count of points within the hypersphere per unit increase of r, across its entire surface

[7] In the absence of a standard implementation of the chi distribution's CDF in the `stats` R package, $P(a, b)$ can be computed using `pgamma(b, a)`.

volume (e.g., sphere surface area if $d = 3$, or circle circumference if $d = 2$). This surface volume is given by

$$s^{(d)}(r) = \frac{d}{dr} v^{(d)}(r) = \frac{\pi^{d/2}\, d\, r^{d-1}}{\Gamma(d/2+1)}. \tag{6.13}$$

Obtaining the expected intensity at a point location that is distance r from the focal point (rather than over the entire surface volume) is therefore achieved by dividing through by $s^{(d)}(r)$; that is,

$$\begin{aligned}
\lambda^{(d)}(r;\boldsymbol{\theta}) &= \left(\frac{d}{dr} n^{(d)}(r;\boldsymbol{\theta})\right) \frac{1}{s^{(d)}(r)} \\
&= \left(\frac{d}{dr} n_d^{(d)}(r; D, \boldsymbol{\psi}) + \frac{d}{dr} n_s^{(d)}(r; \boldsymbol{\psi}, \sigma)\right) \frac{1}{s^{(d)}(r)} \\
&= \left(\frac{d}{dr} D\, E_c(\boldsymbol{\psi})\, v^{(d)}(r) + \frac{d}{dr} [E_s(\boldsymbol{\psi}) - 1]\, F_y^{(d)}(r;\sigma)\right) \frac{1}{s^{(d)}(r)} \\
&= \left(D\, E_c(\boldsymbol{\psi})\, s^{(d)}(r) + [E_s(\boldsymbol{\psi}) - 1]\, f_y^{(d)}(r;\sigma)\right) \frac{1}{s^{(d)}(r)} \\
&= D\, E_c(\boldsymbol{\psi}) + \frac{[E_s(\boldsymbol{\psi}) - 1]\, f_y^{(d)}(r;\sigma)}{s^{(d)}(r)}.
\end{aligned}$$

For increased computational stability and efficiency, the quotient $f_y^{(d)}(r;\sigma)/s^{(d)}(r)$ can be simplified substantially, as this gives

$$\begin{aligned}
\frac{f_y^{(d)}(r;\sigma)}{s^{(d)}(r)} &= \frac{2^{1-d/2}\, r^{d-1}\, \exp(-r^2/(4\sigma^2))}{(\sigma\sqrt{2})^d\, \Gamma(d/2)} \frac{\Gamma(d/2+1)}{\pi^{d/2}\, d\, r^{d-1}} \\
&= \frac{2^{1-d/2}\, \exp(-r^2/(4\sigma^2))\, d}{2\, (\sigma\sqrt{2})^d\, \pi^{d/2}\, d},
\end{aligned}$$

as $\Gamma(d/2+1)/\Gamma(d/2) = d/2$ for all $d \in \mathbb{R}, d > 0$,

$$= \frac{\exp(-r^2/(4\sigma^2))}{(2\,\sigma)^d\, \pi^{d/2}},$$

and, following substitution into the above,

$$\lambda^{(d)}(r;\boldsymbol{\theta}) = D\, E_c(\boldsymbol{\psi}) + \frac{[E_s(\boldsymbol{\psi}) - 1]\, \exp(-r^2/(4\sigma^2))}{(2\,\sigma)^d\, \pi^{d/2}}. \tag{6.14}$$

Without this simplification there is numerical instability in $\lambda^{(d)}(r;\boldsymbol{\theta})$ at $r = 0$, as both the numerator and the denominator in the second term contain r^{d-1}.

Upon substitution of Equations (6.8), (6.11), (6.13), and $d = 2$ into the above, the PI

function presented here simplifies to Equation (6.1), due to Tanaka et al. (2008), for the special two-dimensional case with $C \sim \text{Poisson}(\nu)$.

One important point to note is that the PI function derived here (Equation (6.14)) is fully characterised by just three parameters, these being D, σ, and a *single* parameter from the distribution of C. Recall that the asymptote is controlled by $D\,E_c(\boldsymbol{\psi})$ and the y-intercept by $[E_s(\boldsymbol{\psi}) - 1]/[(2\,\sigma)^d\,\pi^{d/2}]$, while the 'variance' of the Gaussian term is only affected by the parameter σ (Figure 6.2). Say $\boldsymbol{\psi}$ contains two parameters, and that a change in one leads to an increase in both $E_c(\boldsymbol{\psi})$ and $E_s(\boldsymbol{\psi}) - 1$; it is then possible to adjust the other parameter in $\boldsymbol{\psi}$ to return $[E_s(\boldsymbol{\psi}) - 1]$ to its original value. Although $E_c(\boldsymbol{\psi})$ will remain different, a further adjustment to D can ensure that the product $D\,E_c(\boldsymbol{\psi})$ remains unchanged. This new set of parameters has not affected the asymptote or the y-intercept, and so the PI function remains unchanged. As $\lambda^{(d)}(r;\boldsymbol{\theta})$ is not unique for each element of the parameter space, it has become *parameter redundant*, and the full set of parameters cannot be estimated. Thus, identifiability only holds if $\boldsymbol{\psi}$ contains a single parameter.

Further still, the PI function is not only nonidentifiable across the parameter space—but also the model space, as far as the distribution of C is concerned. That is, following a change in the choice of the one-parameter distribution of C, the same PI function can be reconstructed—as long as at least one point in the parameter space of the new distribution gives rise to the same value of $E_s(\boldsymbol{\psi})$.

The Palm likelihood

As per the two-dimensional approach of Tanaka et al. (2008), the d-dimensional difference process is approximated by an inhomogeneous Poisson point process with intensity function decreasing radially from the origin subject to the PI function, $\lambda^{(d)}(r;\boldsymbol{\theta})$.

A generalisation of the Palm likelihood due to Tanaka et al. (2008; Equation (6.2)) is then given by

$$L(\boldsymbol{\theta};\boldsymbol{r}) = \left(\prod_{\{i,j:i\neq j, r_{ij}<t\}} n\,\lambda^{(d)}(r_{ij};\boldsymbol{\theta}) \right) \times \exp\left(-n \int_0^t \lambda^{(d)}(r;\boldsymbol{\theta})\,s^{(d)}(r)\,\mathrm{d}r \right), \qquad (6.15)$$

where the integral in the second term is again a volume of revolution around the intensity axis (here in d-dimensional space), which can be simplified as

$$\int_0^t \lambda^{(d)}(r;\boldsymbol{\theta})\,s^{(d)}(r)\,\mathrm{d}r = \int_0^t \left(D\,E_c(\boldsymbol{\psi}) + \frac{[E_s(\boldsymbol{\psi}) - 1]\,f_y^{(d)}(r;\sigma)}{s^{(d)}(r)} \right) s^{(d)}(r)\,\mathrm{d}r$$

$$
\begin{aligned}
&= \int_0^t D\, E_c(\boldsymbol{\psi})\, s^{(d)}(r) + [E_s(\boldsymbol{\psi}) - 1]\, f_y^{(d)}(r;\sigma)\, \mathrm{d}r \\
&= \left(D\, E_c(\boldsymbol{\psi}) \int_0^t s^{(d)}(r)\, \mathrm{d}r \right) + \left([E_s(\boldsymbol{\psi}) - 1] \int_0^t f_y^{(d)}(r;\sigma)\, \mathrm{d}r \right) \\
&= D\, E_c(\boldsymbol{\psi})\, v^{(d)}(t) + [E_s(\boldsymbol{\psi}) - 1]\, F_y^{(d)}(t;\sigma), \qquad (6.16)
\end{aligned}
$$

and is thus available in closed form. This Palm likelihood is another generalisation of the special case derived by Tanaka et al. (2008).

An estimator for $\boldsymbol{\theta}$ is then

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \boldsymbol{r}), \qquad (6.17)$$

which can be evaluated numerically via maximisation of $\log(L(\boldsymbol{\theta}; \boldsymbol{r}))$, with respect to $\boldsymbol{\theta}$.

Due to the aforementioned nonidentifiability across the model space, an interesting result is that the estimated PI function and maximised Palm likelihood typically remain unchanged despite a change in the modelled distribution of C. The only exception is if one distribution is not flexible enough to obtain an equivalent value for the expected family size of a randomly selected child, $E_s(\boldsymbol{\psi})$. For example, $E_s(\nu) \geq 0$ if $C \sim \text{Poisson}(\nu)$, but a change to $C \sim \text{Binomial}(2, p)$ gives $0 \leq E_s(p) \leq 2$. If the former gives an estimate $E_s(\widehat{\nu}) > 2$, then the two fitted PI functions will be forced to be different. Note that equivalence in fitted PI functions does not imply equivalence in estimated parent densities, $\widehat{D}$—it is the functions of the estimates $\widehat{D}\, E_c(\widehat{\boldsymbol{\psi}})$ and $E_s(\widehat{\boldsymbol{\psi}})$ that remain unchanged, not the parameter estimates themselves.

An unfortunate consequence is that evaluating goodness-of-fit or performing model selection for the distribution fitted to C is not directly possible. Further discussion on this point is presented in Section 6.5.5.

Incorporating known sibling information

In some scenarios information about sibling relationships may be available. Some pairs of observed children may be known siblings and others may be known nonsiblings, leaving only some relationships uncertain (Figure 6.4). Incorporating such information into NSPP parameter estimators has the potential to improve their performance.

In order to accomplish this, the PI function can be partitioned into three components:

$$\lambda^{(d)}(r; \boldsymbol{\theta}) = \lambda_s^{(d)}(r; D, \boldsymbol{\psi}, \alpha) + \lambda_n^{(d)}(r; \boldsymbol{\psi}, \sigma, \beta) + \lambda_u^{(d)}(r; \boldsymbol{\theta}, \alpha, \beta),$$

where (i) $\lambda_s^{(d)}(r; D, \boldsymbol{\psi}, \alpha)$ returns the expected intensity of known siblings, (ii) $\lambda_n^{(d)}(r; \boldsymbol{\psi}, \sigma, \beta)$

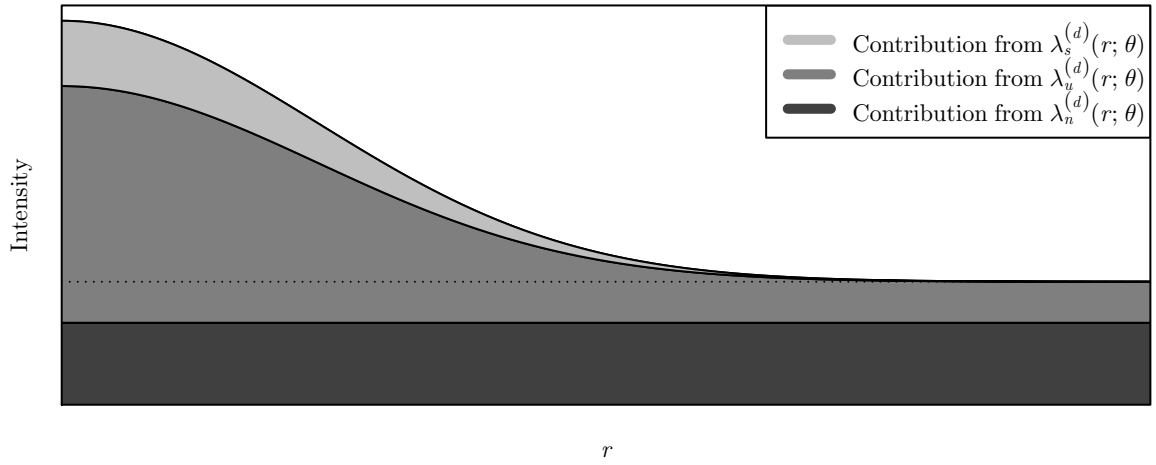


Figure 6.5 *The partitioning of the PI function into $\lambda_s^{(d)}(r;\boldsymbol{\psi},\sigma)$, $\lambda_n^{(d)}(r;d,\boldsymbol{\psi})$, and $\lambda_u^{(d)}(r;\boldsymbol{\theta})$. This plot has been constructed with $\alpha = 1/4$ and $\beta = 2/3$, meaning that one-quarter and two-thirds of the PI function below and above the dotted line are contributions from $\lambda_n^{(d)}(r;d,\boldsymbol{\psi})$ and $\lambda_s^{(d)}(r;\boldsymbol{\psi},\sigma)$, respectively.*

returns the expected intensity of known nonsiblings, and (iii) $\lambda_u^{(d)}(r;\boldsymbol{\theta})$ returns the expected intensity of children that are neither known siblings nor known nonsiblings—all at distance r from a randomly selected, observed point (Figure 6.5).

The constant term $D\,E_c(\boldsymbol{\psi})$ in the PI function (Equation (6.14)) is due to points in the difference process that compare nonsiblings, and the remaining term $\{[E_s(\boldsymbol{\psi}) - 1]\, f_y^{(d)}(r;\sigma)\}/[s^{(d)}(r)]$ is due to points that compare siblings (see Figure 6.2). As the first partition corresponds to siblings that have been successfully identified as such, it accounts for a fraction of the latter term, so that

$$\lambda_s^{(d)}(r;D,\boldsymbol{\psi},\alpha) = \frac{\alpha\,[E_s(\boldsymbol{\psi}) - 1]\, f_y^{(d)}(r;\sigma)}{s^{(d)}(r)}. \tag{6.18}$$

Here α is the probability that two siblings are successfully identified as being siblings. Likewise, the second partition corresponds to nonsiblings that have been successfully identified as such, and accounts for a fraction of the latter term, so that

$$\lambda_n^{(d)}(r;\boldsymbol{\psi},\sigma,\beta) = \beta\,D\,E_c(\boldsymbol{\psi}). \tag{6.19}$$

Here, β is the probability that two nonsiblings are successfully identified as being nonsiblings. In principle it may be possible to estimate α and β from the observed data, but here only scenarios where these are known *a priori* are considered; this is the case for two-plane

cetacean surveys, see Section 6.4.2. The remaining relationships between all pairs of points that have not been identified as either known siblings or known nonsiblings are unknown, and so points in the difference process corresponding to these are incorporated into the final remaining partition, so that

$$\lambda_u^{(d)}(r;\boldsymbol{\theta},\alpha,\beta) = (1-\beta)\,D\,E_c(\boldsymbol{\psi}) + \frac{(1-\alpha)\,[E_s(\boldsymbol{\psi})-1]\,f_y^{(d)}(r;\sigma)}{s^{(d)}(r)}. \tag{6.20}$$

Estimation of $\boldsymbol{\theta}$ can then be achieved by treating the three sets of points in the difference process (i.e., points that compare known siblings, those that compare known nonsiblings, and those that compare points with an unknown relationship) as separate point processes—relying on an inhomogeneous Poisson point process approximation for each. Their corresponding intensity functions depend on Equations (6.18), (6.19), and (6.20), respectively, solely through the distance from the origin. This gives rise to the Palm likelihood

$$\begin{aligned} L(\boldsymbol{\theta};\boldsymbol{r}) = &\left(\prod_{\{i,j:i\neq j,r_{ij}<t,h_{ij}=0\}} n\,\lambda_s^{(d)}(r_{ij};D,\boldsymbol{\psi},\alpha)\right) \times \left(\prod_{\{i,j:i\neq j,r_{ij}<t,h_{ij}=1\}} n\,\lambda_n^{(d)}(r_{ij};\boldsymbol{\psi},\sigma,\beta)\right) \\ &\times \left(\prod_{\{i,j:i\neq j,r_{ij}<t,h_{ij}=2\}} n\,\lambda_u^{(d)}(r_{ij};\boldsymbol{\theta},\alpha,\beta)\right) \times \exp\left(-n\int_0^t \lambda^{(d)}(r;\boldsymbol{\theta})\,s^{(d)}(r)\,\mathrm{d}r\right), \end{aligned} \tag{6.21}$$

where

$$h_{ij} = \begin{cases} 0 & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are known siblings,} \\ 1 & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are known nonsiblings,} \\ 2 & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are neither known siblings nor known nonsiblings.} \end{cases} \tag{6.22}$$

The final multiplicative term of this Palm likelihood remains as per Equation (6.14)—the sums of the volumes of revolution around the intensity axes for each of the three partitions sum to that of the overall Palm intensity (just as the three partitions sum to the overall PI function in Figure 6.5). This is a further generalisation of the Palm likelihood, which collapses to Equation (6.15) when $\alpha = \beta = 0$ (i.e., when no sibling information is known).

6.2.3 Variance estimation

In the absence of a true likelihood, a parametric bootstrap can be used to calculate standard errors and CIs. This is ideal as (i) data simulation is straightforward, and (ii) the point

estimation approach described above is extremely computationally efficient, as the Palm likelihood is available in closed form.

Simulating realisations of NSPPs simply involves (i) generating parent locations with intensity D across a domain that includes $\mathcal{W}$, with an additional buffer of at least $5\,\sigma$ around all edges; (ii) simulating the number of children spawned by each parent from the distribution of C; (iii) dispersing these children around their respective parents, subject to $\mathrm{N}_d(\mathbf{0}_d, \sigma^2\,\boldsymbol{I}_d)$, and (iv) omitting any children that fall outwith $\mathcal{W}$. Step (i) allows for children generated by parents that do not fall within $\mathcal{W}$ to nevertheless be included in the realised point process, thus avoiding edge effects.

Once parameter estimates have been obtained from each bootstrap resample, the same bootstrap variance estimation methods described in Section 2.3.3 can be applied.

6.3 The `nspp` package

The general estimation method described in the previous section has been implemented in the R package `nspp` (Stevenson, 2016). This section introduces, describes, and provides examples for various functions found within the package. All exported data objects and functions have full documentation, and so further information can be found in the package's manual.

6.3.1 The `sim.ns()` function

The `nspp` package comes with two example sets of simulated data—found in the exported objects `example.1D` and `example.2D`—providing realisations of NSPPs in one- and two-dimensional space, respectively. These were generated with `sim.ns()`, a function used to simulate data from NSPPs.

The function has three main arguments:

- `pars`: a vector of model parameters, $\boldsymbol{\theta}$. This must contain named elements `D`, `sigma`, and `child.par`—corresponding to D, σ, and $\boldsymbol{\psi}$, respectively. Recall that $\boldsymbol{\psi}$ must only contain a single parameter.

- `lims`: a matrix with d rows, each giving the limits of a dimension of $\mathcal{W}$. Only (hyper)rectangular observation windows are permitted.

- `rchild`: a function that simulates random variates from the distribution of C. The first argument of this function must provide the number of random variables to gen-

erate (as per standard R functions for random number generation), and the second must provide the parameter $\boldsymbol{\psi}$.

Additional optional parameters allow for the generation of plots of either the simulated point process or the empirical PI function (see Fewster et al., in press). The simulated locations are subject to periodic boundary conditions (PBCs).

For example, simulating data subject to $C \sim \text{Poisson}(\nu)$, $D = 5$, $\sigma = 0.05$, $\nu = 10$, and $\mathcal{W} = [0, 2]^2$ can be achieved using

```
points <- sim.ns(pars = c(D = 5, sigma = 0.05, child.par = 10),
                 lims = rbind(c(0, 2), c(0, 2)), rchild = rpois)
```

The resulting object[8] is a matrix with n rows and 2 columns, providing the locations of the simulated children.

6.3.2 The `fit.ns()` function

Given a set of observed locations, estimation of $\boldsymbol{\theta}$ can be achieved via the `fit.ns()` function. The log of the Palm likelihood (Equation (6.15), or, more generally, (6.21)) is maximised (Equation (6.17)) numerically using a limited-memory, quasi-Newton approach with box constraints (Byrd, Lu, Nocedal, & Zhu, 1995), implemented in the `"L-BFGS-B"` method of the `optim()` function.

A selection of the function's arguments are

- `points`: a matrix containing locations of the observed points. The object returned by `sim.ns()` is suitable.
- `lims`: equivalent to the argument of `sim.ns()` with the same name.
- `R`: the truncation distance t.
- `child.dist`: a list containing information about the distribution fitted to C. Recall that this only enters the likelihood via its expectation, $E_c(\boldsymbol{\psi})$, and variance, $V_c(\boldsymbol{\psi})$. Thus expressions of these in terms of the single parameter $\boldsymbol{\psi}$ must therefore be provided. The required named components of this list are
  - `mean`: a function that takes $\boldsymbol{\psi}$ as its only argument and returns $E_c(\boldsymbol{\psi})$.
  - `var`: a function that takes $\boldsymbol{\psi}$ as its only argument and returns $V_c(\boldsymbol{\psi})$.

[8] Running `set.seed(1234)` prior to executing this code will result in the same `points` object, which is used throughout this section.

  - `sv`: the start value of $\boldsymbol{\psi}$ for the optimisation algorithm.
  - `bounds`: a vector containing bounds for the parameter $\boldsymbol{\psi}$.

- `siblings`: a list containing information about known sibling and known nonsibling relationships. The required named components of this list are
  - `matrix`: a matrix, where the jth element of the ith row is `TRUE` if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are known siblings, `FALSE` if they are known nonsiblings, and `NA` if their relationship is unknown.
  - `pT`: the (fixed) value of α.
  - `pF`: the (fixed) value of β.

Assuming no sibling information is available, $\boldsymbol{\theta}$ can be estimated from the `points` object generated above by `sim.ns()` by running

```
child.dist <- list(mean = function(nu) nu,
                   var = function(nu) nu,
                   sv = 5, bounds = c(0, 100))
fit <- fit.ns(points = points, lims = rbind(c(0, 2), c(0, 2)),
              R = 1, child.dist = child.dist)
```

The resulting object is a list containing various components that provide information about the fitted model. The best way of accessing this information is by extracting it using a variety of utility functions.

6.3.3 Utility functions

A range of S3 methods have been written for generic functions—including `coef()` for parameter estimates, `confint()` for CIs, `plot()` to visualise the fitted PI function, and `summary()` for a model summary. For example:

```
coef(fit)

#           D       sigma   child.par
#  2.14915069  0.06030259 13.68802014

summary(fit)
```

```
# Coefficients:
#            Estimate Std. Error
# D          2.149151
# sigma      0.060303
# child.par 13.688020
```

Notice that `summary()` did not return standard errors—recall that variance estimates are only available following a parametric bootstrap.

6.3.4 The `boot.ns()` function

The relationship between the functions `admbsecr()` and `boot.admbsecr()` in the `admbsecr` package is the same to that between `fit.ns()` and `boot.ns()` here: the object returned by `fit.ns()` can be used as the first argument to `boot.ns()`, and results in the execution of a parametric bootstrap procedure. Again, the argument `N` sets the number of bootstrap resamples to carry out. An additional argument is `rchild`, which provides the function from which the number of children spawned by each parent is simulated. This is equivalent to the argument of `sim.ns()` with the same name.

A parametric bootstrap can therefore be carried out using the function call

```
boot.fit <- boot.ns(fit, rchild = rpois, N = 1000, prog = TRUE)
```

This allows for the display of standard errors when `summary()` is called, and for the calculation of CIs with `confint()`:

```
summary(boot.fit)

# Coefficients:
#            Estimate Std. Error
# D          2.149151     1.0027
# sigma      0.060303     0.0239
# child.par 13.688020     7.8125

confint(boot.fit)

#               2.5 %     97.5 %
# D         0.2830518  4.1419100
# sigma     0.0479578  0.1439433
# child.par 8.7472087 36.2930559
```

6.4 Application to two-plane cetacean surveys

In order to apply the NSPP parameter estimation approach outlined above to data collected from a two-plane cetacean survey, it is necessary to determine (i) the distribution of C, specifically the functions $E_c(\boldsymbol{\psi})$ and $V_c(\boldsymbol{\psi})$ required for $E_s(\boldsymbol{\psi})$ (Equation (6.7)); (ii) known sibling information, h_{ij}, for all pairs of observed cetacean locations; and (iii) values for α and β.

6.4.1 Distribution of child counts

A naïve initial assumption is that each plane sights a given animal independently with probability p, and so $C \sim \text{Binomial}(2, p)$, $\boldsymbol{\psi} \equiv p$. In this case,

$$\begin{aligned} E_s(p) &= \frac{2p\,(1-p) + 4p^2}{2p} - 1 \\ &= p. \end{aligned}$$

However, sightings of a particular individual are unlikely to be independent unless the distance between the planes is substantial (Section 6.1.3); thus, it is preferable to consider the distribution of C to be the sum of two dependent Bernoulli random variables. Let (i) $p_{0|0}$ be the probability that the second plane misses an individual, given that it was missed by the first; (ii) $p_{0|1}$ be the probability that the second plane sights the individual, given that it was missed by the first; (iii) $p_{1|0}$ be the probability that the second plane misses the individual, given that it was sighted by the first; and (iv) $p_{1|1}$ be the probability that the second plane sights the individual, given that it was sighted by the first. Two pairs of these parameters are confounded, as $p_{1|1} = 1 - p_{1|0}$ and $p_{0|0} = 1 - p_{0|1}$. In this context, $p_{1|0}$ and $p_{0|1}$ are considered the free parameters, giving $\boldsymbol{\psi} \equiv (p_{1|0}, p_{0|1})$.

Let the random variable A_i be 1 if a particular individual was sighted by the ith plane and 0 if it was not—so $C = A_1 + A_2$. The random variables $\{A_1, A_2\}$ can be considered a realisation of a two-state Markov chain, with transition probabilities outlined above. The stationary distribution of this Markov chain provides the marginal distribution of A_1 and A_2. Assuming that both planes are equally proficient at detecting individuals (estimation is possible if not, although this is not presented here—see Section 6.4.2), then both PMFs are equivalently

$$f_a(a_i; \boldsymbol{\psi}) = \begin{cases} \frac{p_{1|0}}{p_{1|0}+p_{0|1}} & a_i = 0, \\ \frac{p_{0|1}}{p_{1|0}+p_{0|1}} & a_i = 1, \end{cases}$$

for $i \in \{1, 2\}$. As $A_1, A_2 \sim \text{Bernoulli}(p_{0|1}/p_{1|0} + p_{0|1})$,

$$E_a(\boldsymbol{\psi}) = \frac{p_{0|1}}{p_{1|0} + p_{0|1}} \tag{6.23}$$

and

$$V_a(\boldsymbol{\psi}) = \frac{p_{1|0}\, p_{0|1}}{(p_{1|0} + p_{0|1})^2}. \tag{6.24}$$

Furthermore, the covariance between A_1 and A_2 is

$$\begin{aligned} C_{a_1 a_2}(\boldsymbol{\psi}) &= E_{a_1 a_2}(\boldsymbol{\psi}) - E_{a1}(\boldsymbol{\psi})\, E_{a2}(\boldsymbol{\psi}) \\ &= p_{1|1}\, f_a(1; p_{1|0}, p_{0|1}) - E_a(\boldsymbol{\psi})^2, \end{aligned}$$

as $E_{a_1 a_2}(\boldsymbol{\psi})$ is equivalent to the probability of an individual being sighted by both planes. After substitution and rearrangement, this gives

$$C_{a_1 a_2}(\boldsymbol{\psi}) = \frac{p_{1|0}\, p_{0|1}\, (1 - p_{1|0}^2 - p_{0|1}^2)}{(p_{1|0} + p_{0|1})^2}. \tag{6.25}$$

Equations (6.23), (6.24), and (6.25) provide straightforward derivations for the required $E_c(\boldsymbol{\psi})$ and $V_c(\boldsymbol{\psi})$:

$$\begin{aligned} E_c(\boldsymbol{\psi}) &= 2\, E_a(\boldsymbol{\psi}) \\ &= \frac{2\, p_{1|0}}{p_{1|0} + p_{0|1}}, \text{ and} \qquad (6.26) \\ V_c(\boldsymbol{\psi}) &= 2\, V_a(\boldsymbol{\psi}) + 2\, C_{a_1 a_2}(\boldsymbol{\psi}) \\ &= \frac{2\, p_{1|0}\, p_{0|1}\, (2 - p_{1|0} - p_{0|1})}{(p_{1|0} + p_{0|1})^2}. \qquad (6.27) \end{aligned}$$

One complication here is that $\boldsymbol{\psi}$ contains two parameters, despite the estimation method only being capable of estimating only one for the distribution of C. This is circumvented here by fixing one of the availability parameters ($p_{1|0}$ or $p_{0|1}$) based on a subjective estimate from an expert. In this case the latter is considered fixed, leaving $p_{1|0}$ as the estimable parameter. This issue is not unique to the estimation approach proposed here—Hiby and Lovell (1998) also fixed an availability parameter. It is not yet clear whether the approach due to Borchers et al. (in prep.) suffers the same model identifiability issue.

6.4.2 Sibling information

As per Section 6.1.3 and Figure 6.4, two sightings from the same plane cannot be of the same animal. The formulation of known sibling relationships for two-plane surveys (i.e., a specific form of Equation (6.22)) is therefore

$$h_{ij} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are sightings from the same plane,} \\ 2 & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are sightings from different planes.} \end{cases}$$

Note that h_{ij} cannot be 0 regardless of i and j—two sightings are never confirmed as the same animal; that is, $\alpha = 0$. Additionally, although two sightings made far apart may almost certainly be of different individuals, this information is not incorporated here. The implausibility of a cetacean being sighted in two very different locations is subject to the value of the parameter σ, and so this component of known sibling relationships is handled entirely by the model. Indeed, this estimation approach explicitly aims to avoid any allocation of detections to individuals based solely on the difference between their respective traces.

Recall the assumption that both planes detect individuals equally well. Half of all randomly selected pairs of detections will then correspond to sightings from the same plane, and so half of all comparisons will be between known nonsiblings. Therefore, $\beta = 0.5$. Although not considered here, one plane may be more proficient than the other at making detections. In such a case the relative proficiency may be estimated by observing the relative numbers of detections made—if one plane is twice as proficient, then it should detect approximately twice as many individuals. The estimates $\widehat{\alpha}$ and $\widehat{\beta}$ can then be derived from this estimated relative proficiency.

6.4.3 Implementation using the `nspp` package

Estimating cetacean density from two-plane survey data using the `nspp` package is straightforward. The chosen distribution for C can be correctly specified by creating the following object to use as the `child.dist` argument for `fit.ns()`, as per Equations (6.26) and (6.27):

```
child.dist <- list(mean = function(p10) 2*a/(p10 + a),
                   var = function(p10) 2*p10*a*(2 - p10 - a)/(p10 + a)^2,
                   sv = 0.5,
                   bounds = c(0, 1))
```

Here, `a` must be replaced by the fixed value chosen for $p_{0|1}$.

Creating the matrix required for the `sibling` argument for `fit.ns()` is not straightforward. The `nspp` package therefore includes the function `siblings.twoplane()` to aid in its creation. Its only argument is `plane.id`: a vector containing the plane identification number (e.g., 1 or 2) for each observed point. It then returns the required list, along with the components `pT` and `pF`.

The additional function `sim.twoplane()` can be used for simulation of data from two-plane cetacean surveys. This returns not only the generated observation locations, but also their associated plane identification numbers.

6.4.4 Estimator performance

A simulation study was carried out to evaluate the performance of the proposed estimator (Equation 6.17). Average locations were simulated with homogeneous intensity $D = 1$ cetacean per km over a 100 km one-dimensional transect. The distribution used for cetacean location dispersion around this average location at the time of each plane's overhead pass was $\mathrm{N}(0, \sigma^2 = 0.025^2)$. Positive correlation between sightings across planes was induced by setting $p_{1|0} = 0.1$ and $p_{0|1} = 0.2$. Data were simulated using the `sim.twoplane()` function from the `nspp` R package. In total 10 000 sets of data were generated. The ability to run such a large simulation study is testament to the computational efficiency of the methods described in this chapter.

The models

In Section 6.2 three generalisations of the method due to Tanaka et al. (2008) are made that allow for (i) the modelling of NSPPs in a space of any dimension, (ii) the use of any distribution for C, and (iii) the incorporation of known sibling relationships. In order to investigate the effect of (ii), above, three types of model were fitted to each simulated data set. These are described below.

Tanaka et al. (2008) simply used $E_c(\boldsymbol{\psi})$ in place of $E_s(\boldsymbol{\psi}) - 1$ in Equation (6.5) and claimed this to be appropriate regardless of the distribution of C, showing examples for which $C \sim \mathrm{Poisson}(\nu)$. While the derivation due to Tanaka et al. (2008) is appropriate for this scenario (Equation (6.8)), the result does not hold in general. The exact method proposed by Tanaka et al. (2008; although using the one-dimensional generalisation presented here) is the first model type. This allows for evaluation of robustness of their approach to departures from $C \sim \mathrm{Poisson}(\nu)$.

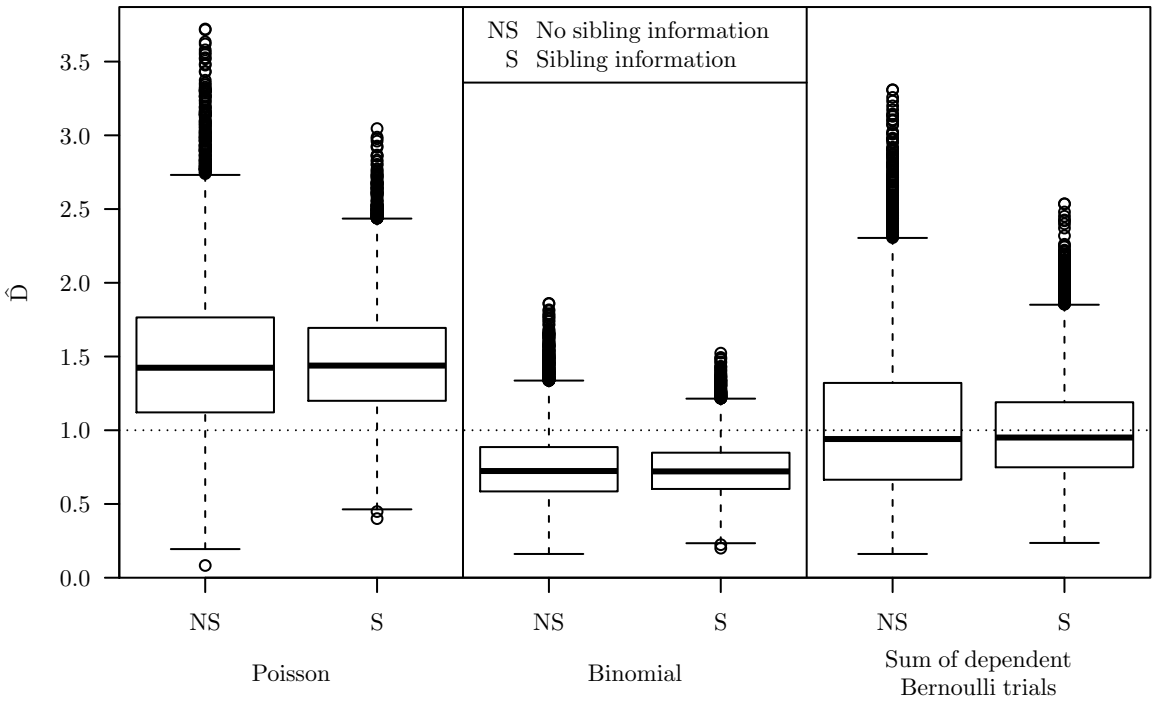


Figure 6.6 *Results from the NSPP simulation study. The dotted horizontal line shows the true density parameter from which the data were generated.*

The second type assumes $C \sim \text{Binomial}(2, p)$, thus representing a model where sightings of an individual are assumed to be independent across planes (an assumption that has previously been made with such data).

The third type assumes the model used to simulate the data; that is, C is a sum of two dependent Bernoulli trials, subject to parameters $p_{1|0}$ and $p_{0|1}$.

In order to investigate the effect of incorporating known sibling relationships, two models were fitted for each model type: (i) a model that ignored the known nonsibling information (i.e., estimators based on maximisation of Equation (6.15)), and (ii) A model that incorporated the known nonsibling information (i.e., estimators based on maximisation of Equation (6.21)).

All models were fitted using the `fit.ns()` function from the `nspp` R package.

Results

Estimated cetacean densities, $\widehat{D}$, across the six models fitted to each simulated set of data are shown in Figure 6.6. Estimated biases, variances and MSEs of the estimators are displayed in Table 6.1. The models that assume C is a sum of dependent Bernoulli trials show negligible bias. Assuming $C \sim \text{Poisson}(\nu)$ and $C \sim \text{Binomial}(p)$ causes positive and

Table 6.1 *Estimated biases, variances, and MSEs from the simulation study for the cetacean density estimators based on the six models investigated. The 'Sibling' column indicates whether or not the known nonsibling information was incorporated.*

Model type	Sibling	Bias	Variance	MSE
Poisson	No	0.465	0.245	0.462
	Yes	0.460	0.139	0.351
Binomial	No	−0.251	0.055	0.118
	Yes	−0.268	0.034	0.106
Dependent Bernoulli	No	0.036	0.231	0.233
	Yes	−0.013	0.104	0.104

negative bias, respectively. For all three model types, incorporation of the known nonsibling information reduces estimator variance.

Interestingly, density estimators from the second model type (based on a binomial assumption for C) have considerably smaller variance than those from the third (based on an assumption that C is a sum of dependent Bernoulli trials) despite the latter being the model from which the data were generated. For models with no plane information incorporated, the variance reduction was so substantial that the density estimator from second model type easily outperformed that of the third (in terms of MSE). This does not imply that correctly modelling C is unimportant and detracts from density estimator performance in general: there is theoretical justification for NSPP estimators that apply to correctly specified models (Prokešová & Vedel Jensen, 2013), and these will not necessarily carry over to misspecified models. Indeed, the second model type shows strong bias and is therefore unlikely to provide a consistent estimator for D (or, at least, bias approaches zero at a slower asymptotic rate than the correctly-specified model). The performance of the third model type relative to the second is likely to improve as the length of the transect surveyed increases. Regardless, once plane information is incorporated, the third model type provides the best-performing cetacean density estimator under the parameter values chosen for data simulation in any case.

6.5 Discussion

This chapter has described three generalisations made to the original estimation method due to Tanaka et al. (2008). Extensions (i) to a spatial domain of any dimension, (ii) to the modelling of any (one-parameter) distribution for C, and (iii) allowing for the incorporation of known sibling (or nonsibling) relationships were all tested in the simulation study

presented above—and performed well on all counts.

The strong bias observed for the first model class (based on a Poisson assumption for C) shows that the method of Tanaka et al. (2008) is not appropriate for any distributional assumption, despite their original claims. The extension made here to use the derived $E_s(\boldsymbol{\psi})-1$ (Equation (6.7)) in place of $E_c(\boldsymbol{\psi})$ in the second term of the PI function (Equation (6.14)) had the desired effect, virtually eliminating estimator bias and reducing variance (Figure 6.6, Table 6.1). Likewise, incorporating the additional sibling information reduced estimator variance; intuitively, the more information an estimation approach uses, the more precise the resulting estimator.

The remainder of this discussion covers improvements to the NSPP estimation procedure, further extensions and applications of TC models beyond NSPPs, and additional modifications that could potentially improve their application to two-plane cetacean surveys.

6.5.1 Analytic estimation

The log of the Palm likelihood is available as a closed-form expression, and, although not shown here, partial derivatives with respect to each of the model parameters can be easily obtained symbolically. Setting these to 0 generates a system of three simultaneous equations. No attempt has yet been made to solve these for the parameters (due to time constraints), but if this is achieved it may give rise to analytic estimation of $\boldsymbol{\theta}$—without the need for any numerical integration or optimisation methods. This is something of a surprising possibility, given how intractable estimation of NSPP parameters appears on the surface. If an analytic solution to these simultaneous equations is not available, then a numerical approach could be employed—whether or not this is more efficient than numerical maximisation of the Palm likelihood is an open question to be evaluated at that juncture.

One advantage of an analytic estimator is a reduction in the (albeit modest) computational burden associated with these methods. For a set computational budget, a larger number of bootstrap resamples could be generated, improving the accuracy of variance estimators. Furthermore, given parameter estimators that are closed-form expressions of the observed data, there is also the possibility of deriving analytic large-sample variance estimators. This represents a substantial second advantage, removing the need for a bootstrap procedure entirely.

In the absence of an existing analytic solution, deriving a heuristic, closed-form estimator for one single parameter as a function of the those remaining is relatively straightforward: the volume beneath the intensity surface (out to a radius of t) of the difference process

(Equation (6.16), multiplied by n), gives the *expected* number of points in the difference process within distance t of the origin, and it is sensible to consider that this should match the corresponding *observed* number, n_r, at the estimate $\widehat{\boldsymbol{\theta}}$. This gives

$$n_r = n\,\{\widehat{D}\,E_c(\widehat{\boldsymbol{\psi}})\,v^{(d)}(t) + [E_s(\widehat{\boldsymbol{\psi}}) - 1]\,F_y^{(d)}(t;\widehat{\sigma})\},$$

and, solving for $\widehat{D}$, for example, gives

$$\widehat{D} = \frac{(n_r/n)\,(E_s(\widehat{\boldsymbol{\psi}}) - 1)\,F_y^{(d)}(t;\widehat{\sigma})}{E_c(\widehat{\boldsymbol{\psi}})\,v^{(d)}(t)}. \tag{6.28}$$

Despite being a heuristic estimator, the above equation has held for all estimates obtained via the maximisation of the Palm likelihood to date, providing empirical evidence to suggest that the two are equivalent. This is advantageous as the optimisation algorithm used to evaluate Equation (6.17) can then be reduced to a search over a two-dimensional parameter space, with the final parameter D being obtained from Equation (6.28). If they cannot be considered equivalent, then Equation (6.28) nevertheless provides an excellent start value for numerical optimisation.

6.5.2 Pseudolikelihood estimation

The formulation of the Palm likelihood has many features similar to that of a *pseudolikelihood* (Besag, 1975, 1977), with a product over pairs of observations that involves some function that compares them. In a pseudolikelihood framework, this function is the distribution of one conditional of the other—here it is the PI function, which is conceptually similar: it is the intensity (rather than probability density) of one, conditional on the location of the other.

Pseudolikelihood is a common framework for parameter estimation of spatial point patterns (e.g., Baddeley & Turner, 2002), and so its use could be considered for TC models. One potential advantage of this is that variance estimation may then be obtained by the 'sandwich formula' approach of White (1982) as an alternative to the bootstrap proposed here.

6.5.3 Extension beyond NSPPs

Here, only Gaussian dispersion of children around parents was considered. TC models can be extended beyond this distributional assumption, however, this requires rederivation of the PDF and CDF of the distance between two randomly selected siblings, $f_y(y;\boldsymbol{\gamma})$ and

$F_y(y;\boldsymbol{\gamma})$.

The only existing example of such a TC model was developed for the application to data from camera-trap detections of *R. rattus* (Fewster et al., in press). As the observed point process was temporal (rather than spatial) in nature, the differences between sibling traces (i.e., times between detections of the same individual) involved a sum of exponential random variables. A further complicating factor was that a detection of an individual triggered the camera to begin recording for a short period of time, and so no further sibling observations could be made during this period. These recording times were also incorporated into the PDF and CDF of intersibling trace differences. The ability to derive the PI and Palm likelihood functions despite this relatively complex dispersion process showcases the great flexibility of TC models, and motivates their potential use in a wide array of ecological scenarios.

6.5.4 Two-plane cetacean surveys

There is scope to further refine the estimation approach described here for two-plane cetacean surveys. The original spatial data collected by the high-definition cameras affixed to the planes are naturally two-dimensional in nature: the resulting images give not only an observed distance along the transect, but also a horizontal distance from its main axis. However, for the methods described here, they must be collapsed over the width of the surveyed region. This leaves only a distance along this main axis. There is information about whale movement—the parameter σ—in these discarded data: the larger σ becomes, the larger the expected intersibling distance across this secondary axis. One such refinement is therefore to retain the information across this second dimension.

Estimation of $\boldsymbol{\theta}$ from the two-dimensional data cannot necessarily be directly achieved using methods developed in this chapter. Recall that the truncation distance, t, must be set at a value that is no greater than half the length of the shortest dimension, in this case, the transect strip width. Given either a narrow enough strip or a large enough plane separation, cetaceans may swim further than than this between the passing of the planes. This truncation will then result in some comparisons between siblings—and almost all comparisons between nonsiblings—being discarded, leaving very little information remaining for parameter estimation.

The narrow nature of the surveyed transects has further implications on the detection process described in Section 6.4.1. The dependence between the random variables A_1 and A_2 was framed as simply being due to the availability of detection due to cetacean diving behaviour. However, an individual is truly only available for detection if it is on the surface

and within the surveyed transect strip. The subjective estimate of $p_{0|1}$ should therefore incorporate both of these components. This is problematic, given that the latter depends on σ, a parameter to be estimated.

An improved estimator should therefore build this second dimension of the observed sighting locations into the derived PI function itself, so that it is not reliant on prohibitive truncation. There is then potential to also separate availability of detection due to diving behaviour (subject to the parameters $p_{1|0}$ and $p_{1|0}$) and location within the surveyed transect strip (subject to the parameter σ). For wide enough transects (or close enough planes), however, a two-dimensional application of the methods presented here may suffice.

These issues are not unique to this estimation approach, and are again shared by that due to Hiby and Lovell (1998), and possibly also Borchers et al. (in prep.). Arguably—as the Palm-likelihood approach is vastly computationally cheaper than its full-likelihood counterparts—there is greater scope to include further complication like a second spatial dimension. Even under the one-dimensional simplification, benchmarking these three approaches against one another is of potential interest. It is possible that the full-likelihood methods provide more information about $\boldsymbol{\theta}$, and nature of the trade-off between estimator performance and computational efficiency is yet to be investigated.

6.5.5 Further extensions

Goodness-of-fit testing for the distribution of C remains problematic, especially as the distributional assumption can substantially alter estimated parent density (e.g., compare the model types in Figure 6.6 and Table 6.1). In some cases, additional information may be available that allows inference into the suitability of the chosen distribution. For example, consider a wildlife survey generating data for a TC model that involves a partially marked animal population. Assuming marked individuals are always successfully identified, C_i is observed if the ith individual is marked. This then would allow for comparison between the fitted and empirical distributions of C.

Furthermore, a Bayesian estimation approach appears to be a natural extension to avoid model nonidentifiability once $\boldsymbol{\psi}$ becomes a vector of parameters. Indeed—for two-plane cetacean surveys—it is far more appropriate to elicit a prior from an expert than a subjective estimate for $p_{1|0}$, which is then considered fixed for estimation purposes. As the latter ignores any variance associated with $p_{1|0}$, variance estimates for all parameters are likely to be negatively biased. Bayesian estimation via the full likelihood using MCMC methods is likely to be infeasible in many cases, as this requires sampling across an extraordinary number of possible animal identity combinations. Once more, implementation via the Palm

likelihood is perhaps the most practical option.

6.5.6 Concluding remarks

Although still in their infancy, TC models appear to be the most practical, efficient, and extensible of those currently available that are capable of dealing with uncertain animal identification in wildlife surveys. While this chapter has provided an array of generalisations specific to parameter estimation from data generated by NSPPs, similar derivations of the PI function for a wide variety of clustered point processes have the potential to model a diverse range of wildlife survey data.

Previous chapters of this thesis have shown the utility of SECR approaches for the modelling of ecological data collected by detectors that passively detect individuals. Issues involving uncertain animal identification are rife in this context. While there is substantial work to be done before SECR and TC approaches can be combined, giving rise to methodology capable of dealing with identification uncertainty within the context of an SECR model, the foundations have nevertheless been laid for future methodological development.

Chapter 7

Discussion

This thesis has primarily focussed on the development of SECR methods, and chiefly those that can be applied to surveys on which animals are detected from their cues. Prior to this methodological development SECR models were not capable of appropriately estimating animal density from any survey that collected cue-based detection data. Now inference is possible given the availability of either independent cue-rate data (Chapter 3) or information about the identity of the animal that produced each detected cue (Chapter 4). Moreover, there are now methods to account for situations in which cues vary in detectability or have a directional component (Chapter 5). In each of the above cases the collection of auxiliary data is either required, or can substantially improve estimator performance. The incorporation of such data into SECR models is available through the methods developed in Chapter 2.

There is thus scope for the development of further methods capable of estimating animal density from surveys on which neither cue-rate nor identity data are available. The class of TC models described in Chapter 6 shows promise in accounting for uncertain animal identification not just for SECR models, but their use also extends throughout statistical ecology—and even beyond. For example, TC models implemented in the `nspp` package are currently being used for the classification of malignant tumours through analysis of clustered nuclei locations at the cellular level (C. M. Jones-Todd, personal communication, 16 December 2015).

A common theme in the discussion appearing across many chapters (see Sections 3.6.6, 4.6.5, and 5.5.6) is the need to develop methods that require information about individual ID and are capable of estimating cue-production rates from the capture histories themselves, without the need for the collection of independent data. Furthermore, the methods described in Chapters 3 and 4 both come with the fundamental assumption that animals

remain stationary throughout the survey. This is appropriate for the applications to *A. lightfooti* and *S. aurocapilla* detection data presented therein, but this does not hold for cue-detection surveys of many species.

The following sections in this chapter focus chiefly on methods to solve these salient issues in cue-based analysis of SECR data. Section 7.1 broaches the issue of estimating cue-production rates in addition to other SECR parameters (e.g., animal density, the detection function, and auxiliary information measurement error) from only the detection data collected by the main survey. The majority of this section, in fact, is not devoted to discussion at all—instead, a new likelihood is derived that shows promise in achieving this aim. This is given particular attention here as it appears to be a means of incorporating a range of extensions made within this thesis into a single SECR model. It has the potential to provide a flexible framework under which analysis of cue detection data can be analysed. Section 7.2 broaches issues involving animal movement. In this case a likelihood is not derived, but potential means of achieving this are discussed.

7.1 Integrated cue-rate estimation from SECR surveys

The main difficulty associated with estimation of cue production rates from capture histories is that only some cues from each individual are detected, and it is not known how many cues produced by each individual were missed. Any method that attempts to achieve this must somehow account for cues that have been missed by the detector array.

In this section a likelihood is proposed, maximisation of which has the potential to estimate the mean population call-production rate (and therefore animal density) from a cue-based SECR survey. It is assumed that each detected cue can be attributed to an individual (see Table 1.1a). This requires slight modification to notation set out earlier:

1. Let n_a be the total number of unique individuals detected.

2. Let n_i be the number of detected cues that were produced by the ith detected individual, and let this be the ith element of the vector $\boldsymbol{n}$. Let n^* be the number of detected cues that were produced by a randomly selected individual in the population. There is a subtle distinction between the distributions of n_i and n^*: the variable n_i is only observed for individuals that had at least one cue detected. Therefore, $n_i \geq 1$, and its distribution is a truncated version of that associated with n^*.

3. Let r_i be the *total* number of calls that were produced by the ith detected individual over the course of the survey. Let r be the total number of cues that were produced

by a randomly selected individual in the population. Its PDF is characterised by parameters held in the vector $\boldsymbol{\zeta}$.

4. Let $\omega_{ijk} = 1$ if the kth detector detected the jth detected cue produced by the ith detected individual, and $\omega_{ijk} = 0$ otherwise. The vector $\boldsymbol{\omega}_{ij}$ now provides the capture history corresponding to the jth detected cue that was produced by the ith detected individual, and $\boldsymbol{\Omega}_i$ is now a matrix containing capture histories associated with detected calls produced by the ith detected individual.

5. Likewise, let $\boldsymbol{y}_{ijk}$ be the vector of auxiliary information collected by the kth detector on the jth detected cue that was produced by the ith detected individual; this may contain, for example, the types of data discussed in Chapter 2. As above, the matrix $\boldsymbol{Y}_{ij}$ contains auxiliary information collected on the jth detected call that was produced by the ith detected individual, and the array $\boldsymbol{Y}_i$ contains all auxiliary data related to all detections of calls produced by the ith detected individual.

The total number of cues produced by the ith individual, r_i is not directly observed; only the number of detected cues is, and so it is only known that $r_i \geq n_i$.

The likelihood is the joint density of the observed data as a function of the model parameters; this is therefore given by

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{n}) &= f(\boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{n}, n_a; \boldsymbol{\theta}) \\ &= f(n_a; \boldsymbol{\theta})\, f(\boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{n} | \boldsymbol{n}_a; \boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta}). \end{aligned}$$

Assuming independence between individuals,

$$= f(n_a; \boldsymbol{\theta}) \prod_{i=1}^{n_a} f(\boldsymbol{\Omega}_i, \boldsymbol{Y}_i, n_i; \boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta}).$$

The location of the ith individual affects its capture histories, observed auxiliary information, and its number of detected cues, and so integration over $\boldsymbol{x}_i$ is required, giving

$$\begin{aligned} &= f(n_a; \boldsymbol{\theta}) \prod_{i=1}^{n_a} \int_{\mathcal{A}} f(\boldsymbol{\Omega}_i, \boldsymbol{Y}_i, n_i | \boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{x}_i. \\ &= f(n_a; \boldsymbol{\theta}) \prod_{i=1}^{n_a} \int_{\mathcal{A}} f(\boldsymbol{\Omega}_i, \boldsymbol{Y}_i | \boldsymbol{x}_i, n_i; \boldsymbol{\gamma}, \boldsymbol{\psi})\, f(n_i | \boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{x}_i. \end{aligned}$$

Assuming independent detections between cues produced by the same individual (conditional on its location) gives

$$= f(n_a;\boldsymbol{\theta}) \prod_{i=1}^{n_a} \int_{\mathcal{A}} \left(\prod_{j=1}^{n_i} f(\boldsymbol{\omega}_{ij}, \boldsymbol{Y}_{ij} | \boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\psi}) \right) f(n_i | \boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})\, f(\boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{x}_i. \tag{7.1}$$

Each of the terms appearing above is now derived in turn.

The term appearing in the product over j is the joint PDF of a detected cue's capture history and its associated observed auxiliary information, conditional on the location it was produced. This can be be rewritten as

$$f(\boldsymbol{\omega}_{ij}, \boldsymbol{Y}_{ij} | \boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\psi}) = f(\boldsymbol{Y}_{ij} | \boldsymbol{\omega}_{ij}, \boldsymbol{x}_i; \boldsymbol{\psi})\, f(\boldsymbol{\omega}_{ij} | \boldsymbol{x}_i; \boldsymbol{\psi}), \tag{7.2}$$

both terms of which have already been derived in Chapters 1 and 2 for a range of auxiliary data types; see Equation (1.8) for the latter and Equations (2.4) and (2.6) for the former, and specifically Equation (2.7) for estimated distances, (2.9) for estimated bearings, (2.12) for received signal strengths, and (2.14) for received TOAs. Additionally, CSH can be incorporated by replacing Equation (5.19) from Chapter 5 with the above Equation (7.2). CD can also be incorporated but this is not shown here.

For derivation of the remaining terms, many functions defined previously for SECR methodology depend on the number of cues produced by an individual. For example, the probability of detecting an individual located at $\boldsymbol{x}$ increases with r; the more cues an individual produces the more likely it is to be detected. Let the subscript r indicate that a function is conditional on the number of cues produced by an individual; for example, $p_r(\boldsymbol{x}|r; \boldsymbol{\gamma}, \boldsymbol{\zeta})$ returns the probability of an individual at $\boldsymbol{x}$ being detected given that it produced r cues.

As per Chapter 4, let $p_c(\boldsymbol{x}; \boldsymbol{\gamma})$ be the probability of a particular call produced at $\boldsymbol{x}$ being detected by at least one detector. This remains

$$p_c(\boldsymbol{x}; \boldsymbol{\gamma}) = 1 - \prod_{k=1}^{m} [1 - g(d_k(\boldsymbol{x}); \boldsymbol{\gamma})];$$

see Equations (1.3) and (4.2). Assuming independence across cues produced by the same individual,

$$n^* | \boldsymbol{x}, r \sim \text{Binomial}(r, p_c(\boldsymbol{x}; \boldsymbol{\gamma})), \tag{7.3}$$

and this gives

$$f(n_i^*|\boldsymbol{x}_i, r_i; \boldsymbol{\gamma}, \boldsymbol{\zeta}) = \binom{r_i}{n_i^*} p_c(\boldsymbol{x}; \boldsymbol{\gamma})^{n_i^*} \left(1 - p_c(\boldsymbol{x}; \boldsymbol{\gamma})\right)^{r_i - n_i^*}. \tag{7.4}$$

The number of cues produced by the ith individual, r_i, is a latent variable, and so to calculate the PMF of n_i conditional only on $\boldsymbol{x}_i$ (rather than on both $\boldsymbol{x}_i$ and r_i) it is necessary to sum over all possible values r could take. This provides the second term in the integrand of (7.1):

$$f(n_i|\boldsymbol{x}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta}) = \frac{\sum_{r_i=n_i}^{\infty} f(n_i^*|\boldsymbol{x}_i, r_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})\, f(r_i; \boldsymbol{\zeta})}{p(\boldsymbol{x}; \boldsymbol{\gamma}, \boldsymbol{\zeta})}. \tag{7.5}$$

The lower limit of this sum can be set at n_i rather than 0, as the summand is 0 if $r_i < n_i$; the number of detected calls cannot exceed the number of total calls. The first term in this summand is the binomial PMF (as per Equation (7.4)), while the choice of the second is a modelling decision. If it can be assumed that cue production is a Poisson process with the same rate parameter across all individuals, then $f(r_i; \boldsymbol{\zeta})$ is the Poisson PMF, in which case $\boldsymbol{\zeta}$ simply contains the expected total number of calls each individual makes over the duration of the survey. If there is variance in cue production rates across the population, then it may be sensible to consider this to be the negative binomial PMF[1]. The denominator of Equation (7.5) exists by virtue of never being able to observe $n_i = 0$; it is not known when an individual is not detected at all, and so n_i has a truncated binomial distribution.

This denominator is the probability of detecting at least one cue from an individual located at $\boldsymbol{x}$, and is therefore given by the DPS, $p(\boldsymbol{x}; \boldsymbol{\gamma}, \boldsymbol{\zeta})$. It is derived here for cases where individuals produce multiple cues. First, the probability of detecting an individual located at $\boldsymbol{x}$ given that it produced r cues is the probability that at least one of its r cues was detected by at least one detector, and so

$$\begin{aligned} p_r(\boldsymbol{x}|r; \boldsymbol{\gamma}) &= \Pr(n^* \geq 1|\boldsymbol{x}, r) \\ &= 1 - f_{n^*|\boldsymbol{x}r}(0|\boldsymbol{x}, r; \boldsymbol{\gamma}), \end{aligned}$$

where $f_{n^*|\boldsymbol{x}r}(\cdot)$ is given by Equation (7.4). Therefore,

$$p_r(\boldsymbol{x}|r; \boldsymbol{\gamma}) = 1 - \left[1 - p_c(\boldsymbol{x}; \boldsymbol{\gamma})\right]^r.$$

[1]The negative binomial distribution can be considered a mixture, whereby the Poisson rate parameter for each observation is a realisation of a random variable with a gamma distribution.

To obtain the marginal probability of detection it is then necessary to sum over all possible values r may take, as before:

$$p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{\zeta}) = \sum_{r=0}^{\infty} p_r(\boldsymbol{x}|r;\boldsymbol{\gamma})\, f(r;\boldsymbol{\zeta}). \tag{7.6}$$

Note that this includes the case $r = 0$, and thus remains accurate even in cases where some individuals do not call[2]. This completes derivation of Equation (7.5) and also aids in that of the remaining two terms in the likelihood, Equation (7.1): both remain almost equivalent to definitions provided in Chapter 1. The marginal PDF for a randomly selected detected individual remains

$$f(\boldsymbol{x}_i;\boldsymbol{\gamma},\boldsymbol{\zeta}) = \frac{p(\boldsymbol{x}_i;\boldsymbol{\gamma},\boldsymbol{\zeta})}{\int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{x}},$$

as per Equation (1.4), and the number of unique individuals remains a Poisson random variable with expectation

$$\mu = \mathrm{E}(n_a) = D\int_{\mathcal{A}} p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{x},$$

as per Equation (1.7); however, in both cases, the rederived $p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{\zeta})$ (Equation (7.6)) is now used. Despite the data consisting of cue-based capture histories, the parameter D nevertheless refers to *animal* density, as $p(\boldsymbol{x};\boldsymbol{\gamma},\boldsymbol{\zeta})$ now also refers to the probability of detecting individuals.

Estimation of $\boldsymbol{\theta} = (D,\boldsymbol{\gamma},\boldsymbol{\psi},\boldsymbol{\zeta})$ can then once again be achieved via maximisation of the likelihood function, Equation (7.1). Estimation under a Bayesian framework is an attractive alternative due to the complex nature of the latent animal locations and cue frequencies. In this case, each $\boldsymbol{x}_i$ and r_i could be sampled at every iteration of an MCMC algorithm, removing the need for the direct evaluation of the likelihood function with marginalisation over these latent variables.

At the time of writing, no attempt has yet been made to implement this method in any software package. Its usefulness therefore remains uncertain, but—at face value—it appears a promising means of analysing cue-based data collected by an array of fixed detectors.

In fact, it has the potential to extend quite far beyond this setting—for example, to general situations in which many units (of any kind) can be detected that are associated in some way with the same point in space. Consider a camera-trap SECR survey in which each detected individual belongs to a family, and all members of a family have the same home-

[2]This holds only if $\Pr(r = 0)$ is appropriately estimated once $\widehat{\boldsymbol{\zeta}}$ is obtained. If r has a zero-inflated distribution, for example, then this is unlikely.

range centre (e.g., by virtue of sharing a burrow or nest). In such a case there are strong analogies between the two situations. Rather than $\boldsymbol{\omega}_{ij}$ being the capture history associated with the jth detected cue produced by the ith detected individual, consider it to be that associated with the jth detected member of the ith detected family. Rather than $\boldsymbol{x}_i$ being the physical location of the ith detected individual, consider it to be the location of the ith detected family's home-range centre. Rather than r_i being the number of detectable cues produced by the ith individual, consider it to be the number of detectable individuals in the ith family. In other words, the approach described here has the potential to model the dependence between locations associated with groups of detection units—be they individual cues or individual animals—and this is a feature that is almost always ignored in the fitting of SECR models.

Furthermore, specification of this likelihood for situations where identities are observed may provide a stepping stone towards deriving estimators for situations in which they are not, thus accounting for identification uncertainty in SECR surveys. For example, animal density estimation in such a case is potentially possible by summing this likelihood over all possible combinations of identity allocations to detected cues—analogous to the approaches of Borchers et al. (in prep.) and Hiby and Lovell (1998) for the analysis of two-plane survey data discussed in Chapter 6. Granted, aside from the most trivial cases, a naïve implementation of this approach is almost certainly computationally intractable.

7.2 Animal movement in cue-based SECR surveys

All methods developed throughout Chapters 2–5, along with that proposed in the previous section, rely on the assumption that cue-producing individuals do not move throughout the duration of the survey. This allows them to be associated with just one single location within the survey region. While this is appropriate for many anurans (e.g., *A. lightfooti*) and some songbirds (e.g., *S. aurocapilla*), there is current interest in the use of cue-based SECR approaches to estimate densities of species belonging to taxa for which this does not hold. Examples include cetaceans (e.g., *B. acutorostrata*, *P. macrocephalus*, the harbour porpoise *Phocoena phocoena*) and microbats (e.g., the northern long-eared bat *Myotis septentrionalis*). Methods that account for intercue movement of individuals are required in order to relax this assumption and therefore obtain appropriate estimates of animal density in these cases. However, it is worth noting that (ironically) the more animals move between the production of cues, the less of a concern this becomes; for example, if intercue movement is so great that the location at which one cue is produced does not provide any information

about the next, then these locations can be considered independent. The methods that suitably estimate animal density with stationary animals can then instead be directly used to estimate cue density.

Standard DS methods also assume that individuals remain stationary. Animal movement can cause substantial bias in density estimates obtained from both point-transect and line-transect surveys, particularly at higher animal speeds (Glennie, Buckland, & Thomas, 2015; Prieto, Thomas, & Marques, in prep.), and so this is also possibly the case for current cue-based SECR estimates. DS methodology to account for movement on such surveys, however, does not yet exist.

SECR methods for dealing with issues pertaining to the movement of animal home-range centres across sampling occasions have recently been proposed (Ergon & Gardner, 2014; Royle, Fuller, & Sutherland, 2016). Modelled movement behaviours include *transience* (whereby individuals' home-range centres move between each sampling occasion) and *dispersal* (whereby individuals may be associated with the same home-range centre for a number of sampling occasions, before undertaking a single movement). The spatial scale of dispersal is usually greater than that of transience. Interestingly, Royle et al. (2016) found (via simulation) that models that did not account for these movement effects nevertheless estimated density with minimal bias.

On SECR surveys for which these movement behaviours have been developed, each capture history is associated with a particular individual and a particular sampling occasion. On cue-based SECR surveys, on the other hand, each capture history is associated with a particular individual and, specifically, a particular *cue* it has produced. Once this comparison between sampling occasions and cues has been made, similarities between movement of home-range centres in standard SECR and movement of individuals in cue-based SECR are obvious: for example, echolocating species typically move around in between the emission of cues (cf. transience), while others (e.g., songbirds) may produce a number of cues at one location before moving to another (cf. dispersal). Indeed, similar models to those that have been proposed to account for transience and dispersal could potentially be used for intercue movement in cue-based analyses.

There is one considerable complication that precludes the direct application of these models to cue-based data. On standard SECR surveys, if a previously detected individual is not detected on one particular occasion, then a capture history of $\mathbf{0}_m$ is observed; thus it is known when a previously detected individual has evaded detection. On cue-based SECR surveys, however, it is not known when a produced cue remains undetected, and so a capture history is *not* observed when a previously detected individual has evaded detection.

This is problematic as it is difficult to determine the plausible range of intercue movement when not all cues generate observed data. For example, if a considerable length of time has elapsed between the detections of two cues attributed to the same individual, then it is plausible that either (i) it produced a number of cues that remained undetected by virtue of having moved away from the detector array, or (ii) it remained in the proximity of the detector array, but did not produce any cues in the intervening period.

In order for a ML approach to infer which is more likely given the data at hand, it is thus necessary not only to integrate over the latent locations at which the two detected cues were produced—but additionally *all possible paths* the individual may have travelled between them, presenting a substantial computational obstacle. This can potentially be minimised by assuming a Markovian movement model—the resulting dependence structure between an individual's locations over the course of a survey allows a high-dimensional integral to be broken down into a series of lower-dimensional integrals, hence increasing efficiency; see Rue and Held (2005). Alternatively, a Bayesian approach could instead sample over plausible latent locations and their connecting paths.

The likelihood derived in the previous section (Equation (7.1)) provides a means of incorporating animal movement. The final term within the integrand needs to be extended to the joint PDF $f(\boldsymbol{X}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta})$, where $\boldsymbol{X}_i$ holds the locations at which all n_i detected cues were produced. The integral for the ith detected individual must be evaluated jointly over all such locations and thus has dimension $2n_i$. The term within the product should then condition on $\boldsymbol{x}_{ij}$, the location at which the jth detected cue was produced by the ith detected individual. Calculation of the PMF of the random variable $n_i|\boldsymbol{X}_i$ remains unclear, and depends on the individual's travelled path—the longer this has the individual in the proximity of the detector array, the larger its expectation.

7.3 Concluding remarks

Two of the most pertinent extensions to cue-based SECR methods that remain unexplored have been discussed over the course of the last two sections; in particular, the likelihood derived for integrated cue-rate estimation appears the most general and extensible yet—it seems to be capable of simultaneously incorporating the use of auxiliary data (Chapter 2), cue rate estimation (without independent cue-rate data collection, as per Chapter 3), cue directionality or strength heterogeneity (Chapter 5), and animal movement (Section 7.2, above).

One particular advancement that requires specific attention—and this is applicable to

SECR methods in general—is the development of further means to assess goodness of fit. Aside from a Monte Carlo test based on the scaled model deviance (Borchers & Efford, 2008), these are virtually nonexistent. While SECR animal density point estimators appear remarkably robust to violations of various model assumptions (e.g., Distiller & Borchers, 2015; Efford & Fewster, 2013; Efford & Mowat, 2014; Royle et al., 2016; Stevenson et al., 2015), this is certainly not always the case (e.g., Chapter 5), and—at present—there are very few ways to identify when such violations have occurred.

This thesis has described novel SECR methods for cue-based detection data that can be appropriately applied to real data. Prior to this work, in order to estimate animal density practitioners were forced to either (i) assume that all individuals produced exactly one cue throughout the duration of the survey, or (ii) restrict analysis to only the first detected cue produced by a handful of individuals, and then proceed as usual. The former will quite clearly almost never hold, while the latter is now discouraged in favour of the work presented in Chapter 4.

In particular, estimation of animal abundance and density from acoustic survey data is becoming increasingly common (Marques et al., 2013); however, there appears to be a dearth of methods capable of analysing detection data collected by fixed arrays of acoustic detectors—especially when the locations of sound sources cannot be deduced with negligible uncertainty. Cue-based SECR methods seem well-suited to fulfil this niche. With the ongoing technological advancement in devices capable of collecting such data, their use is becoming both cheaper and more widespread; the continued development of SECR methodology will conceivably become instrumental in the assessment of various animal populations.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *2nd international symposium on information theory* (pp. 267–281). Budapest: Akadémia Kiadó.

Baddeley, A., & Turner, R. (2002). Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*, *42*, 283–322.

Baudin, M. (1981). Likelihood and nearest-neighbor distance properties of multidimensional Poisson cluster processes. *Journal of Applied Probability*, *18*, 879–888.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *24*, 179–195.

Besag, J. (1977). Efficiency of pseudolikelihood estimation for simple gaussian fields. *Biometrika*, *64*, 616–618.

Blocker, A. W. (2014). fastGHQuad: Fast Rcpp implementation of Gauss-Hermite quadrature [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=fastGHQuad` (R package version 0.2)

Borchers, D. L., Buckland, S. T., & Zucchini, W. (2002). *Estimating animal abundance: Closed populations*. London: Springer.

Borchers, D. L., & Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, *64*, 377–385.

Borchers, D. L., & Langrock, R. (2015). Double-observer line transect surveys with Markov-modulated Poisson process models for animal availability. *Biometrics*, *71*, 1060–1069.

Borchers, D. L., Nightingale, P., & Fewster, R. M. (in prep.). Unmarked recapture methods for digital aerial surveys of marine mammals.

Borchers, D. L., Stevenson, B. C., Kidney, D., Thomas, L., & Marques, T. A. (2015). A unifying model for capture-recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association*, *110*, 195–204.

Borchers, D. L., Zucchini, W., & Fewster, R. M. (1998). Mark-recapture models for line transect surveys. *Biometrics*, *54*, 1207–1220.

Buckland, S. T. (2006). Point-transect surveys for songbirds. *The Auk*, *123*, 345–357.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling: Estimating abundance of biological populations*. Oxford: Oxford University Press.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2004). *Advanced distance sampling: Estimating abundance of biological populations*. Oxford: Oxford University Press.

Buckland, S. T., Oedekoven, C. S., & Borchers, D. L. (2016). Model-based distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, *21*, 58–75.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*, 1190–1208.

Cavanaugh, J. E., & Shumway, R. H. (1997). A bootstrap variant of AIC for state-space model selection. *Statistica Sinica*, *7*, 473–496.

Chandler, R. B., & Royle, J. A. (2012). Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*, *7*, 936–954.

Channing, A. (2004). Genus *Arthroleptella*. In L. R. Minter, M. Burger, J. A. Harrison, H. H. Braack, P. J. Bishop, & D. Kloepfer (Eds.), *Atlas and red data book of the frogs of South Africa, Lesotho and Swaziland* (pp. 206–219). Washington D.C.: Smithsonian Institute.

Cormack, R. M., & Jupp, P. E. (1991). Inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika*, *78*, 911–916.

Crainiceanu, C. M., & Ruppert, D. (2003). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B*, *66*, 165–185.

Curnow, R. N., & Dunnett, C. W. (1962). The numerical evaluation of certain multivariate normal integrals. *The Annals of Mathematical Statistics*, *33*, 571–579.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Dawson, D. K., & Efford, M. G. (2009). Bird population density estimated from acoustic signals. *Journal of Applied Ecology*, *46*, 1201–1209.

Dawson, S. M., Slooten, E., DuFresne, S., Wade, P., & Clement, D. (2004). Small-boat surveys for coastal dolphins: Line transect surveys for Hector's dolphins

(*Cephalorhynchus hectori*). *Fishery Bulletin*, *102*, 441–451.

Distiller, G., & Borchers, D. L. (2015). A spatially explicit capture-recapture estimator for single-catch traps. *Ecology and Evolution*, *5*, 5075–5087.

Dorcas, M. E., Price, S. J., Walls, S. C., & Barichivich, W. J. (2009). Auditory monitoring of anuran populations. In C. K. Dodd (Ed.), *Amphibian ecology and conservation: A handbook of techniques* (pp. 281–298). New York: Oxford University Press.

Dunnett, C. W., & Sobel, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika*, *42*, 258–260.

Eaton, M. L. (1983). *Multivariate statistics: A vector space approach*. New York: Wiley.

Efford, M. G. (2004). Density estimation in live-trapping studies. *Oikos*, *106*, 598–610.

Efford, M. G. (2015). secr: Spatially explicit capture-recapture models [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=secr` (R package version 2.9.4)

Efford, M. G., Borchers, D. L., & Byrom, A. E. (2009). Density estimation by spatially explicit capture-recapture: Likelihood-based methods. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling deomgraphic processes in marked populations* (pp. 255–269). New York: Springer.

Efford, M. G., Dawson, D. K., & Borchers, D. L. (2009). Population density estimated from locations of individuals on a passive detector array. *Ecology*, *90*, 2676–2682.

Efford, M. G., & Fewster, R. M. (2013). Estimating population size by spatially explicit capture-recapture. *Oikos*, *122*, 918–928.

Efford, M. G., & Mowat, G. (2014). Compensatory heterogeneity in spatially explicit capture-recapture data. *Ecology*, *95*, 1341–1348.

Ergon, T., & Gardner, B. (2014). Separating mortality and emigration: Modelling space use, dispersal and survival with robust-design spatial capture-recapture data. *Methods in Ecology and Evolution*, *5*, 1327–1336.

Fewster, R. M., & Jupp, P. E. (2013). Information on parameters of interest decreases under transformations. *Journal of Multivariate Analysis*, *120*, 34–39.

Fewster, R. M., Stevenson, B. C., & Borchers, D. L. (in press). Trace-contrast models for capture-recapture without capture histories. *Statistical Science*.

Foote, J. R., Palazzi, E., & Mennill, D. J. (2013). Songs of the eastern phoebe, a suboscine songbird, are individually distinctive but do not vary geographically. *Bioacoustics*, *22*, 137–151.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., ...

Sibert, J. (2012). AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*, 233–249.

Gagneux, O., Boesch, C., & Woodruff, D. S. (1997). Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology*, *6*, 861–868.

Geweke, J. (1991). Efficient simulation from the multivariate and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface* (pp. 571–578). Fairfax: Interface Foundation of North America, Inc.

Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D., Redmond, P., ... Deng, X. Y. (2008). PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans. *Proceedings of the Institute of Acoustics*, *30*.

Glennie, R., Buckland, S. T., & Thomas, L. (2015). The effect of animal movement on line transect estimates of abundance. *PLoS ONE*, *10*.

Graunt, J. (1662). *Natural and political observations made upon the bills of mortality.* London.

Griewank, A., & Walther, A. (2002). *Evaluating derivatives: Principles and techniques of algorithmic differentiation.* Philadelphia: Society for Industrial and Applied Mathematics.

Hauselberger, K. F., & Alford, R. A. (2005). Effects of season and weather on calling in the Australian mycrohylid frogs *Austrochaperina robusta* and *Cophixalus ornatus*. *Herpetologica*, *61*, 349–363.

Hiby, A. R. (1985). An approach to estimating population densities of great whales from sighting surveys. *IMA Journal of Mathematics Applied in Medicine and Biology*, *2*, 201–220.

Hiby, L., & Lovell, P. (1998). Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Biometrics*, *54*, 1280–1289.

Hiby, L., Paterson, W. D., Redman, P., Watkins, J., Twiss, S. D., & Pomeroy, P. (2012). Analysis of photo-id data allowing for missed matches and individuals identified from opposite sides. *Methods in Ecology and Evolution*, *4*, 252–259.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Hutchinson, J. M. C., & Waser, P. M. (2007). Use, misuse and extensions of "ideal gas"

models of animal encounter. *Biological Reviews*, *82*, 335–359.

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester: Wiley.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions*. New York: Wiley.

Kidney, D., Rawson, B., Borchers, D. L., Marques, T. A., Thomas, L., & Stevenson, B. C. (in submission). An efficient acoustic density estimation method with human detectors, applied to gibbons in Cambodia. *PLoS ONE*.

Kirlin, M. S., Gooch, M. M., Price, S. J., & Dorcas, M. E. (2006). Predictors of winter anuran calling activity in the North Caroline piedmont. *Journal of the North Carolina Academy of Science*, *122*, 10–18.

Koehler, E., Brown, E., & Haneuse, S. J.-P. A. (2009). On the assessment of Monte Carlo error in simulation-based statistical analyses. *The American Statistician*, *63*, 155–162.

Koornwinder, T. H., Roderick, R. S. C., Koekoek, R., & Swarttouw, R. F. (2010). Orthogonal polynomials. In F. W. J. Olver, D. M. Lozier, R. F. Boisvert, & C. W. Clark (Eds.), *NIST handbook of mathematical functions* (pp. 435–484). New York: Cambridge University Press.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). *Continuous multivariate distributions, Volume 1: Models and applications* (2nd ed.). New York: Wiley.

Kraaij, T., & van Wilgen, B. W. (2014). Drivers, ecology, and management of fire in fynbos. In N. Allsopp, J. F. Colville, & G. A. Verboom (Eds.), *Fynbos: Ecology, evolution, and conservation of a megadiverse region* (pp. 47–72). Oxford: Oxford University Press.

Laiolo, P., Vögeli, M., Serrano, D., & Tella, J. L. (2007). Testing acoustic vs physical marking: Two complementary methods for individual-based monitoring of elusive species. *Journal of Avian Biology*, *38*, 672–681.

Laplace, P. S. (1786). Sur les naissances, les mariages et les morts. *Mémoires de l'Acadeémie Royale des Sciences de Paris, Année 1783*, 693–702.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, *59*, 1123–1130.

Link, W. A., Yoshizaki, J., Bailey, L. L., & Pollock, K. H. (2010). Uncovering a latent multinomial: Analysis of mark-recapture data with misidentification. *Biometrics*, *66*, 178–185.

Lucas, T. C. D., Moorcroft, E. A., Freeman, R., Rowcliffe, J. M., & Jones, K. E. (2015). A

generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution*, *6*, 500–509.

Manly, B. F. J., McDonald, L. L., & Garner, G. W. (1996). Maximum likelihood estimation for the double-count method with independent observers. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*, 170–189.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Jarvis, S., Morrissey, R. P., ... DiMarzio, N. (2012). Spatially explicit capture-recapture methods to estimate minke whale density from data collected at bottom-mounted hydrophones. *Journal of Ornithology*, *152*, 445–455.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Morettie, D. J., ... Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, *88*, 287–309.

Martin, S. W., Marques, T. A., Thomas, L., Morrissey, R. P., Jarvis, S., DiMarzio, N., ... Mellinger, D. K. (2013). Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors. *Marine Mammal Science*, *29*, 142–158.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, *36*, 318–324.

Measey, G. J. (2011). Appendix 1. In G. J. Measey (Ed.), *Ensuring a future for South Africa's frogs: A strategy for conservation research.* SANBI Biodiversity Series 19 (pp. 42–83). Pretoria: South African National Biodiversity Institute.

Measey, G. J., Stevenson, B. C., Scott, T., Altwegg, R., & Borchers, D. L. (in submission). Counting chirps: Acoustic monitoring of cryptic frog species. *Journal of Applied Ecology*.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., & Surlykke, A. (2000). Sperm whale clicks: Directionality and source level revisited. *Journal of the Acoustical Society of America*, *107*, 638–648.

Morrison, T. A., Yoshizaki, J., Nichols, J. D., & Bolger, D. T. (2011). Estimating survival in photographic capture-recapture studies: Overcoming misidentification error. *Methods in Ecology and Evolution*, *2*, 454–463.

Murphy, C. G. (2003). The cause of correlations between nightly number of male and female barking treefrogs (*Hyla gratiosa*) attending choruses. *Behavioral Ecology*, *14*, 274–281.

Navas, C. A. (1996). The effect of temperature on the vocal activity of tropical anurans: A comparison of high and low-eleveation species. *Journal of Herpetology*, *30*, 488–497.

Neyman, J., & Scott, E. L. (1952). A theory of the spatial distribution of galaxies. *The*

Astrophysical Journal, *116*, 144–163.

Obbard, M. E., Howe, E. J., & Kyle, C. J. (2010). Empirical comparison of density estimators for large carnivores. *Journal of Applied Ecology*, *47*, 76–84.

Okamura, H., Minamikawa, S., Skaug, H. J., & Kishiro, T. (2012). Abundance estimation of long-diving animals using line transect methods. *Biometrics*, *68*, 504–513.

Oseen, K. L., & Wassersug, R. J. (2002). Environmental factors influencing calling in sympatric anurans. *Oecologia*, *133*, 616–625.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, *62*, 3–135.

Patricelli, G. L., Dantzker, M. S., & Bradbury, J. W. (2007). Differences in acoustic directionality among vocalizations of the male red-winged blackbird (*Agelaius pheoniceus*) are related to function in communication. *Behavioral Ecology and Sociobiology*, *61*, 1099–1110.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545–554.

Petrusková, T., Pišvejcová, I., Kinštová, A., Brinke, T., & Petrusek, A. (2016). Repertoire-based individual acoustic monitoring of a migratory passerine bird with complex song as an efficient tool for tracking territorial dynamics and annual return rates. *Methods in Ecology and Evolution*, *7*, 274–284.

Pike, D. G., Paxton, C. G. M., Gunnlaugsson, T., & Vikingsson, G. A. (2009). Trends in the distribution and abundance of cetaceans from aerial surveys in Icelandic coastal waters, 1986–2001. *NAMMCO Scientific Publications*, *7*, 117–142.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, *56*, 434–442.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge: Cambridge University Press.

Prieto, R., Thomas, L., & Marques, T. A. (in prep.). The effect of animal movement on point transect estimates of abundance.

Prokešová, M., & Vedel Jensen, E. B. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *65*, 387–412.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rowcliffe, J. M., Field, J., Turvey, S. T., & Carbone, C. (2008). Estimating animal density

using camera traps without the need for individual recognition. *Journal of Applied Ecology*, *45*, 1228–1236.

Royle, J. A., Fuller, A. K., & Sutherland, C. (2016). Spatial capture-recapture models allowing Markovian transience or dispersal. *Population Ecology*, *58*, 53–62.

Royle, J. A., Karanty, K. U., Gopalaswamy, A. M., & Kumar, N. S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, *90*, 3233–3244.

Royle, J. A., & Young, K. V. (2008). A hierarchical model for spatial capture-recapture data. *Ecology*, *89*, 2281–2289.

Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and application.* London: Chapman and Hall—CRC Press.

Saenz, D., Fitzgerald, L. A., Baum, K. A., & Conner, R. N. (2006). Abiotic correlates of anuran calling phenology: The importance of rain, temperature, and season. *Herpetological Monographs*, *20*, 64–82.

Schofield, M. R., & Bonner, S. J. (2015). Connecting the latent multinomial. *Biometrics*, *71*, 1070–1080.

Schweder, T., Skaug, H. J., Langaas, M., & Dimakos, X. K. (1999). Simulated likelihood methods for complex double-platform line transect surveys. *Biometrics*, *55*, 678–687.

Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters.* New York: MacMillan Press.

Slooten, E., & Dawson, S. M. (1994). Hector's dolphin. In S. H. Ridgway & R. Harrison (Eds.), *Handbook of marine mammals, Volume V (Delphinidae and Phocoenidae)* (pp. 311–333). New York: Academic Press.

Slooten, E., Dawson, S. M., & Rayment, W. J. (2004). Aerial surveys for coastal dolphins: Abundance of Hector's dolphins off the South Island West Coast, New Zealand. *Marine Mammal Science*, *20*, 477–490.

Stevenson, B. C. (2016). Estimation of parameters for Neyman-Scott point processes [Computer software manual]. Retrieved from `https://github.com/b-steve/nspp` (R package version 0.0.1)

Stevenson, B. C., & Borchers, D. L. (2015). SECR models with supplementary location information [Computer software manual]. Retrieved from `https://github.com/b-steve/admbsecr` (R package version 1.2.0)

Stevenson, B. C., Borchers, D. L., Altwegg, R., Swift, R. J., Gillespie, D. M., & Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, *6*, 38–48.

Stowell, D., & Plumbley, M. D. (2014). Large-scale analysis of frequency modulation in birdsong data bases. *Methods in Ecology and Evolution*, *5*, 901–912.

Sutherland, C., Fuller, A. K., & Royle, J. A. (2014). Modelling non-Euclidean movement and landscape connectivity in highly structured ecological networks. *Methods in Ecology and Evolution*, *6*, 169–177.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., ... Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*, 3189–3194.

Tanaka, U., Ogata, Y., & Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, *50*, 43–57.

Tancredi, A., Auger-Méthé, M., Marcoux, M., & Liseo, B. (2013). Accounting for matching uncertainty in two stage capture-recapture experiments using photographic measurements of natural marks. *Environmental and Ecological Statistics*, *20*, 647–665.

Trautmann, H., Steuer, D., Mersmann, O., & Bornkamp, B. (2014). truncnorm: Truncated normal distribution [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=truncnorm` (R package version 1.0-7)

Turner, A. A., & Channing, A. (2008). A new species of *arthroleptella* Hewitt, 1926 (Anura: Pyxicephalidae) from the Klein Swartberg mountain, Caledon, South Africa. *African Journal of Herpetology*, *57*, 1–12.

Turner, A. A., De Villiers, A. L., Dawood, A., & Channing, A. (2004). A new species of *Arthroleptella* Hewitt, 1926 (Anura: Ranidae) from the Groot Winterhoek mountains of the Western Cape province, South Africa. *African Journal of Herpetology*, *53*, 1–12.

Vale, R. T. R., Fewster, R. M., Carroll, E. L., & Patenaude, N. J. (2014). Maximum likelihood estimation for model $M_{t,\alpha}$ for capture-recapture data with misidentification. *Biometrics*, *70*, 962–971.

Varadhan, R., & Borchers, H. W. (2011). dfoptim: Derivative-free optimization [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=dfoptim` (R package version 2011.8-1)

Weir, L. A., Royle, J. A., & Jung, R. E. (2005). Modeling anuran detection and site occupancy in North American amphibian monitoring program (NAAMP) routes in Maryland. *Journal of Herpetology*, *39*, 627–639.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25.

Wright, J. A., Barker, R. J., Schofield, M. R., Frantz, A. C., Byrom, A. E., & Gleeson,

M. (2009). Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics*, *65*, 833–840.

Xia, C. W., Huang, R., Wei, C., Nie, P., & Zhang, Y. (2001). Individual identification on the basis of the songs of the Asian stubtail (*Urosphena squameiceps*). *Chinese Birds*, *2*, 132–139.

Zimmer, W. M., Tyack, P. L., Johnson, M. P., & Madsen, P. T. (2005). Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. *Journal of the Acoustical Society of America*, *117*, 1473–1485.

Acronyms

AD automatic differentiation
ADMB AD model builder
AIC Akaike information criterion
BIC Bayesian information criterion
ARS automated recording system
CD cue directionality
CDF cumulative distribution function
CI confidence interval
CR capture-recapture
CRAN the Comprehensive R Archive Network
CSH cue strength heterogeneity
CV coefficient of variation
DPS detection probability surface
DS distance sampling
EDF empirical distribution function
ESA effective sampling area
IID independent and identically distributed
MCE Monte Carlo error
MCMC Markov chain Monte Carlo
MCS manual calling survey
ML maximum-likelihood
MRDS mark-recapture distance sampling
MSE mean-squared error
NSPP Neyman Scott point process
PBC periodic boundary condition
PDF probability density function
PI Palm intensity
PMF probability mass function
QQ quantile-quantile
REM random encounter model
RMSE root mean-squared error
SECR spatially explicit capture-recapture
TC trace-contrast
TOA time of arrival

Names of species

Arthroleptella lightfooti	Cape Peninsula moss frog
Balaenoptera acutorostrata	common minke whale
Nomascus annamensis	northern yellow-cheeked gibbon
Myotis septentrionalis	northern long-eared bat
Physeter macrocephalus	sperm whale
Phocoena phocoena	harbour porpoise
Rattus rattus	ship rat
Seiurus aurocapilla	ovenbird

Notation (Chapters 1–5)

Survey layout

m	The number of detectors deployed.
q	The number of observed auxiliary information types collected for each detection.
t	The duration of the survey.
$\mathcal{A}$	The set of all locations in the survey region at which a detected unit may be located.
$\mathcal{M}$	The set of all habitat mask points; $\mathcal{M} \subset \mathcal{A}$.
b	The habitat mask's buffer distance. No points in the mask are beyond distance b from the nearest detector.
$\boldsymbol{x}$	A vector containing the Cartesian coordinates of a location in the survey region at which a detected unit may be located; $\boldsymbol{x} \in \mathcal{A}$.
s	The speed of sound. This only needs to be known for surveys using TOA auxiliary information, and depends on the medium through which acoustic cues propagate.
c	The threshold signal strength for detection of acoustic signals.
c_l	The lower threshold of detection for first-call models.
c_u	The upper threshold of inclusion for first-call models.
$\boldsymbol{c}$	A vector containing both lower and upper thresholds; $\boldsymbol{c} = (c_l, c_u)$.

Observed data

n The number of detected units.

ω_{ij} A detection indicator. If the jth detector detected the ith detected unit, then $\omega_{ij} = 1$, otherwise $\omega_{ij} = 0$.

$\boldsymbol{\omega}_i$ A vector containing the observed capture history of the ith detected unit; $\boldsymbol{\omega}_i = (\omega_{i1}, \cdots, \omega_{im})$.

$\boldsymbol{\Omega}$ A matrix containing the capture histories of all detected units; $\boldsymbol{\Omega} = (\boldsymbol{\omega}_i, \cdots, \boldsymbol{\omega}_n)$.

z_i The number of detectors that detected the ith unit; $z_i = \sum_{j=1}^{m} \boldsymbol{\omega}_{ij}$.

y_{ijk} The kth type of observed auxiliary information, measured from the detection of the ith detected unit at the jth detector.

$\boldsymbol{y}_{ij}$ A vector containing all types of auxiliary information, measured from the detection of the ith detected unit at the jth detector.

$\boldsymbol{y}_{i\cdot k}$ A vector containing the k^{th} type of observed auxiliary information, measured from the detection of the ith detected across all detectors.

$\boldsymbol{Y}_i$ A matrix containing all observed auxiliary information types, measured from the detection of the ith detected unit across all detectors; $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}, \cdots, \boldsymbol{y}_{im})$.

$\boldsymbol{Y}$ A matrix containing all observed auxiliary information; $\boldsymbol{Y} = (\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n)$.

n_r The number of individuals from which call-rate data are collected.

$\boldsymbol{r}$ A vector containing call rates from all n_r monitored individuals.

r_i The call rate observed from the ith monitored individual.

Latent variables

$\boldsymbol{x}_i$ A vector containing the Cartesian coordinates of the (unobserved) location of the ith detected unit; $\boldsymbol{x}_i \in \mathcal{A}$.

$\boldsymbol{X}$ A matrix containing the Cartesian coordinates of the (unobserved) locations of all n detected units; $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, $\boldsymbol{X} \in \mathcal{A}^n$.

u_i The bearing in which the ith detected cue was emitted; $u_i \in [0, 2\pi)$.

v_i The source signal strength of the ith detected cue.

e_{ij} The measurement error in the measured received signal strength of the ith detected cue at the jth detector.

Parameters

D The density parameter. This is cue density if detection units are cues, and animal density if detection units are either individuals or first calls. Named `D` in `admbsecr`.

D_a Animal density (if cues are the unit of detection). Named `Da` in `admbsecr`.

$\boldsymbol{\gamma}$ A vector containing detection function parameters.

$\boldsymbol{\psi}$ A vector containing parameters that characterise the distributions of the auxiliary information types.

α A parameter controlling the measurement error of estimated distances. Named `alpha` in `admbsecr`.

κ A parameter controlling the measurement error of estimated bearings. Named `kappa` in `admbsecr`.

β_0 The source strength of an acoustic signal, measured on the scale of some link function. Named `beta0.ss` in `admbsecr`.

β_1 The received signal strength loss per metre separation between a detector and the signal source, measured on the scale of some link function. Named `beta1.ss` in `admbsecr`.

σ_s A parameter controlling the measurement error of recorded signal strengths. Named `sigma.ss` in `admbsecr`.

β_2 A parameter controlling cue directionality. Specifically, it gives the difference between the signal strength loss rate between the direction the cue was emitted and the direction opposite to this. Named `beta2.ss` in `admbsecr`.

σ_v The population standard deviation of cue source signal strengths. Named `sigma.b0.ss` in `admbsecr`.

$\boldsymbol{\Sigma}$ The variance-covariance matrix for received signal strengths across the detectors.

σ_t A parameter controlling the measurement error of recorded TOAs. Named `sigma.toa` in `admbsecr`.

μ_r The population's mean call rate.

$\boldsymbol{\zeta}$ A vector of parameters for the distribution of population call rates. Thus, $\mu_r \in \boldsymbol{\zeta}$, or μ_r is some function of $\boldsymbol{\zeta}$.

$\boldsymbol{\theta}$ A vector containing all model parameters, i.e., $D, \boldsymbol{\gamma}, \boldsymbol{\psi} \subset \boldsymbol{\theta}$. This also comprises $\boldsymbol{\zeta}$ for surveys that collect independent call-rate data.

Functions

$f(\cdot \mid \cdot\,; \cdot)$ The (joint) PDF or PMF of its argument(s). If a conditional PDF or PMF, the variables it is conditional on appear following a vertical line. Constant(s) (e.g., parameters) are separated from the argument(s) by a semicolon.

$F(\cdot \mid \cdot\,; \cdot)$ The (joint) CDF of its argument(s). If a conditional CDF, the variables it is conditional on appear following a vertical line. Constant(s) (e.g., parameters) are separated from the argument(s) by a semicolon.

$\phi(\cdot)$ The PDF of the standard normal distribution.

$\Phi(\cdot)$ The CDF of the standard normal distribution.

$L(\cdot\,; \cdot)$ A likelihood (or likelihood-type) function. Observed data are separated from the arguments by a semicolon.

$g(\cdot\,; \cdot)$ A detection function, returning the probability of detection of a unit at a detector, with the distance between them specified by the argument. Constant(s) (e.g., parameters) are separated from this argument by a semicolon.

$p(\cdot\,;\cdot)$ A DPS function, returning the probability that a unit located at the location defined by the argument is detected by at least one detector. Constant(s) (e.g., parameters) are separated from this argument by a semicolon. In Chapter 4, this is the probability of an individual being *included*, rather than a call being *detected*.

$p_c(\cdot\,;\cdot, c)$ The probability of a call emitted at the location defined by the argument being detected at a strength louder than some constant c. Constant(s) (e.g., parameters) are separated from this argument by a semicolon. Only used in Chapter 4.

$p_f(\cdot\,;\cdot)$ The probability of an individual at the location defined by the argument being included due to detection of its first call. Constant(s) (e.g., parameters) are separated from this argument by a semicolon. Only used in Chapter 4.

$p_s(\cdot\,;\cdot)$ The probability of an individual at the location defined by the argument being included due to detection of a subsequent call. Constant(s) (e.g., parameters) are separated from this argument by a semicolon. Only used in Chapter 4.

$a(\cdot)$ A function that takes the detection parameters as an argument, and returns the ESA.

$d_j(\cdot)$ The distance between the jth detector and the location defined by the argument.

$\boldsymbol{d}(\cdot)$ A vector containing distances between the location defined by the argument and all m detectors; $\boldsymbol{d}(\boldsymbol{x}) = (d_1(\boldsymbol{x}), \cdots, d_m(\boldsymbol{x}))$.

$d_{\cdot}(\cdot)$ The distance between the centroid of the detectors and the location defined by the argument.

$b_j(\cdot)$ The bearing from the jth detector to the location defined by the argument.

$l_j(\cdot|\cdot;\cdot)$ The signal strength loss rate in the direction of the jth detector. Only used in models incorporating cue directionality.

$\mathrm{erf}(\cdot)$ The error function; $\mathrm{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2}\,\mathrm{d}t$.

Notation (Chapter 6)

Statistical notation

X	An uppercase letter denotes a random variable.
x	A lowercase letter denotes a fixed value that corresponds to the associated random variable in some way (e.g., an observed value or a quantile from its PDF).
$f_x(x;\boldsymbol{\theta})$	The PDF of the random variable X, characterised by the parameter vector $\boldsymbol{\theta}$.
$F_x(x;\boldsymbol{\theta})$	The CDF of the random variable X, characterised by the parameter vector $\boldsymbol{\theta}$.
$E_x(\boldsymbol{\theta})$	The expectation of the random variable X, i.e., $E_x(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} x\, f_x(x;\boldsymbol{\theta})\,\mathrm{d}x$.
$V_x(\boldsymbol{\theta})$	The variance of the random variable X, i.e., $V_x(\boldsymbol{\theta}) = E_{x^2}(\boldsymbol{\theta}) - E_x(\boldsymbol{\theta})^2$.
$C_{xy}(\boldsymbol{\theta},\boldsymbol{\gamma})$	The covariance of the random variables X and Y, where $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ collectively contain their parameters.

TC model notation

d	The dimensionality of the point process.
$\mathcal{W}$	The observation window; $\mathcal{W} \subset \mathbb{R}^d$.
n	The number of observed children.
m	The number of unobserved parents.
$\boldsymbol{x}_i$	A vector containing the coordinates of the location of the ith observed child.
$\boldsymbol{d}$	A vector containing the coordinates at which an untruncated point in the distance process may be located.

$\boldsymbol{d}_{ij}$ The difference between the locations of the ith and jth observed children, i.e., $\boldsymbol{d}_{ij} = (x_{j1} - x_{i1}, \cdots, x_{jd} - x_{id})$.

t The threshold distance from the origin beyond which points in the difference process are truncated.

n_r The number of points in the difference process within distance t of the origin. Equivalently, the number of pairs of observed children within distance t of one another.

r The distance between the origin and a location at which a point in the difference process may be located, i.e., $r = \|\boldsymbol{d}\|$. Equivalently, a possible distance between two children, subject to PBCs.

r_{ij} The distance between the origin and the point in the difference process due to the comparison of the ith and jth children, i.e., $r_{ij} = \|\boldsymbol{d}_{ij}\|$. Equivalently, the distance between the ith and jth children, subject to PBCs.

y The distance between two randomly selected siblings.

c The number of children generated by a randomly selected parent.

c_i The number of children generated by the ith parent.

s The number of siblings owned by a randomly chosen child. This does not include itself.

α The probability that two siblings are successfully identified as such.

β The probability that two non-siblings are successfully identified as such.

$\boldsymbol{r}$ A vector containing distances between the origin and all points in the difference process. Thus, this vector comprises all elements in the set $\{r_{ij} : i, j \in \{1, \cdots, n\}, i \neq j\}$.

a_i An indicator for cetacean detection, where $a = 1$ if the ith plane detected a particular individual, and $a = 0$ if it did not.

h_{ij} An indicator for the status of the known relationship between the ith and jth observed points. If they are known siblings, then $h_{ij} = 0$; if they are known non-siblings, then $h_{ij} = 1$; and if their relationship remains unknown then $h_{ij} = 2$.

Parameters

D The density of parents.

ν The expectation for the number of children generated by a randomly selected parent. This only applies to NSPPs where C is a Poisson random variable, i.e, $C \sim \text{Poisson}(\nu)$.

$p_{\cdot\cdot}$ Depending on the subscripts, the probability of the second plane sighting/missing a whale given that it was sighted/missed by the first plane. For two-plane cetacean surveys.

σ The square-root of the diagonal elements of $\boldsymbol{\Sigma}$, the variance-covariance matrix for the Gaussian dispersion of children around parents. This only applies to isotropic NSPPs where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_d$.

$\boldsymbol{\psi}$ A vector of parameters that characterises the distribution of C. Thus, if C is a Poisson random variable, $\boldsymbol{\psi} \equiv \nu$.

$\boldsymbol{\gamma}$ A vector of parameters that characterises the distribution of the dispersion of children around parents. Thus, if this is Gaussian with variance-covariance matrix of the form $\boldsymbol{\Sigma} = \sigma \boldsymbol{I}_d$, then $\boldsymbol{\gamma} \equiv \sigma$.

$\boldsymbol{\theta}$ A vector containing all model parameters, i.e., $\boldsymbol{\theta} = (D, \boldsymbol{\psi}, \boldsymbol{\gamma})$.

Functions

$\lambda(r; \boldsymbol{\theta})$ The Palm intensity function. This returns the expected point density at a location distance r from an observed child.

$\lambda_s(r; \boldsymbol{\theta})$ The Palm intensity function for comparisons between known siblings. This returns the expected density of known siblings at a location distance r from an observed child.

$\lambda_n(r; \boldsymbol{\theta})$ The Palm intensity function for comparisons between known non-siblings. This returns the expected density of known non-siblings at a location distance r from an observed child.

$\lambda_u(r;\boldsymbol{\theta})$ The Palm intensity function for comparisons between children whose sibling status is unknown. This returns the expected density of points that are neither known siblings nor known non-siblings at a location distance r from an observed child.

$n^{(d)}(r;\boldsymbol{\theta})$ The expected number of other children within distance r of an observed child.

$n_d^{(d)}(r;\boldsymbol{\theta})$ The expected number of children within distance r of an observed child that have a different parent to this focal point.

$n_s^{(d)}(r;\boldsymbol{\theta})$ The expected number of other children within distance r of an observed child that have the same parent as this focal point.

$v^{(d)}(r)$ The volume of the d-dimensional hypersphere with radius r.

$s^{(d)}(r)$ The volume of the surface of the d-dimensional hypersphere with radius r. If $d = 2$, this is the circumference of a circle, and if $d = 3$, this is the surface area of a sphere.